

LB 164

CRFE

Access DB# 79000

# SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: M.A. WALICKA Examiner #: 78201 Date: Oct. 29 '02  
 Art Unit: 1652 Phone Number 305-7270 Serial Number: 09/724,571  
 Mail Box and Bldg/Room Location: 10D06 Results Format Preferred (circle): PAPER DISK E-MAIL  
10D01

If more than one search is submitted, please prioritize searches in order of need. me

\*\*\*\*\*  
 Please provide a detailed statement of the search topic; and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Beta-secretase enzyme composition and method  
 Inventors (please provide full names): Anderson

Earliest Priority Filing Date: 06/15/1999

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search:

SEQ ID NO: 75

103 } these are polypeptides  
 104 } of 4 AA  
 72 } these are polypeptides  
 78 } of 8-15 AA  
 81 }  
 91 }  
 103-361  
 104-4  
 72-14  
 78-7  
 81-9  
 91-9

Thank you in advance.

M. Walicka

Edward Hant  
 Technical Info. Specialist  
 STIC/Biotech  
 CMI 6B02 Tel: 305-9203

RECEIVED  
 OCT 30  
 2002

## STAFF USE ONLY

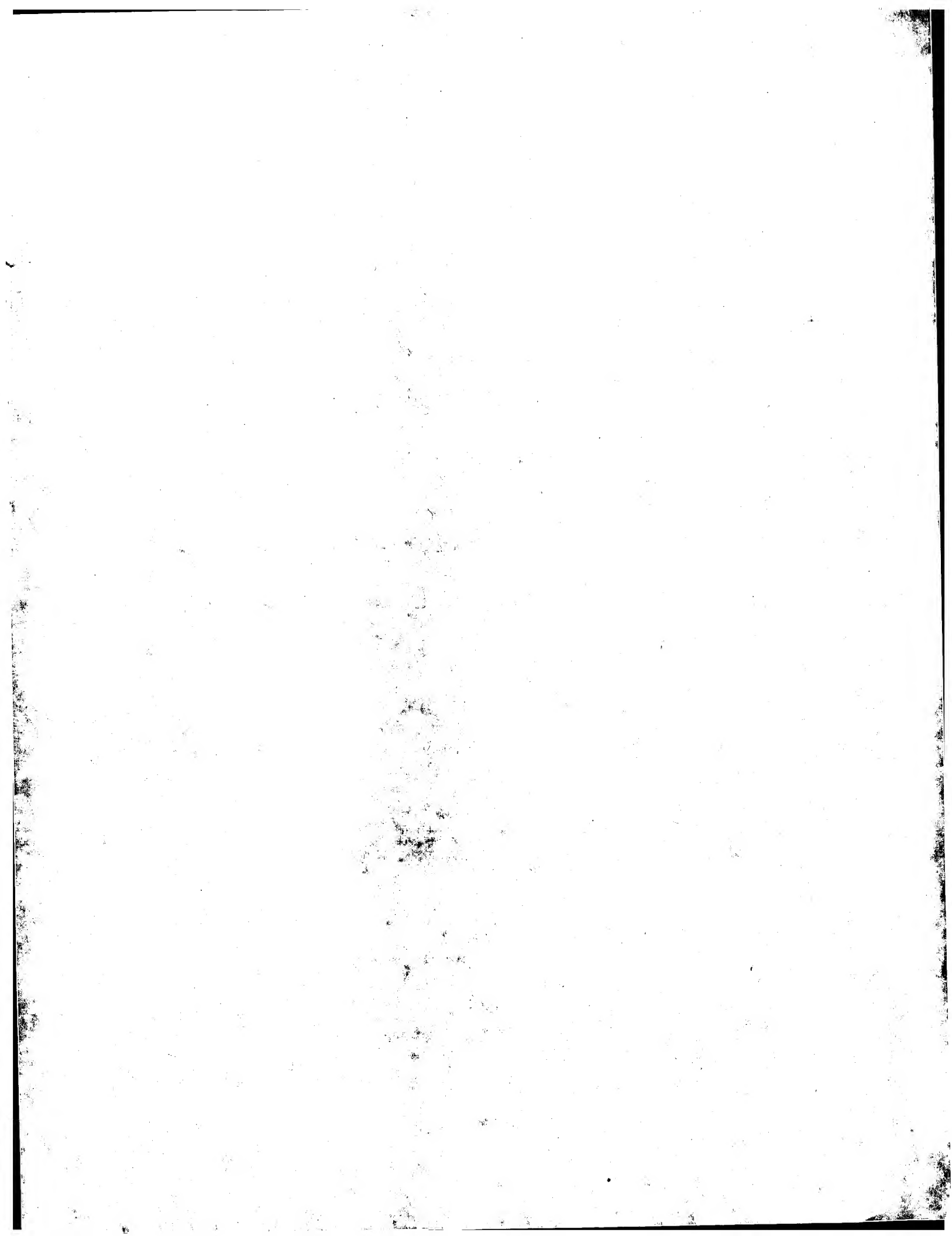
Searcher: \_\_\_\_\_  
 Searcher Phone #: \_\_\_\_\_  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: 10/30/02  
 Date Completed: 10/30/02  
 Searcher Prep & Review Time: \_\_\_\_\_  
 Clerical Prep Time: \_\_\_\_\_  
 Online Time: \_\_\_\_\_

### Type of Search

NA Sequence (#) \_\_\_\_\_  
 AA Sequence (#) 7  
 Structure (#) \_\_\_\_\_  
 Bibliographic \_\_\_\_\_  
 Litigation \_\_\_\_\_  
 Fulltext \_\_\_\_\_  
 Patent Family \_\_\_\_\_  
 Other \_\_\_\_\_

### Vendors and cost where applicable

STN \_\_\_\_\_  
 Dialog \_\_\_\_\_  
 Questel/Orbit \_\_\_\_\_  
 Dr. Link \_\_\_\_\_  
 Lexis/Nexis \_\_\_\_\_  
 Sequence Systems 02  
 WWW/Internet \_\_\_\_\_  
 Other (specify) \_\_\_\_\_





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:20:50 ; Search time 1.60197 Seconds  
(without alignments)  
277.344 Million cell updates/sec

Title: us-09-724-571-104

Perfect score: 20  
Sequence: 1 VNLD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : \_AGeneSeq\_032802.\*

1: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	21	AA1980.DAT.*
2	20	100.0	5	18	AA1981.DAT.*
3	20	100.0	5	19	AA1982.DAT.*
4	20	100.0	5	20	AA1983.DAT.*
5	20	100.0	5	22	AA1984.DAT.*
6	20	100.0	5	21	AA1985.DAT.*
7	20	100.0	8	22	AA1986.DAT.*
8	20	100.0	8	22	AA1987.DAT.*
9	20	100.0	8	22	AA1988.DAT.*
10	20	100.0	9	19	AA1989.DAT.*
11	20	100.0	9	19	AA1990.DAT.*

12	20	100.0	9	21	AA1991.DAT.*	A peptide fragment
13	20	100.0	9	21	AA1992.DAT.*	Substrate for beta
14	20	100.0	9	22	AA1993.DAT.*	Protease inhibitor
15	20	100.0	10	18	AA1994.DAT.*	Beta-secretase sub
16	20	100.0	10	20	AA1995.DAT.*	Synthetic oligopep
17	20	100.0	10	21	AA1996.DAT.*	Beta-APP alpha-sec
18	20	100.0	10	22	AA1997.DAT.*	Human APP-Sw beta
19	20	100.0	10	22	AA1998.DAT.*	Human amyloid prec
20	20	100.0	10	22	AA1999.DAT.*	Synthetic Asp2 rec
21	20	100.0	10	22	AA2000.DAT.*	Human beta-amyloid
22	20	100.0	10	22	AA2001.DAT.*	Human APP-Sw beta
23	20	100.0	10	22	AA2002.DAT.*	Peptide 5-5' SW, fo
24	20	100.0	10	22	AA2003.DAT.*	Synthetic peptide
25	20	100.0	10	22	AA2004.DAT.*	Swedish mutation p
26	20	100.0	11	22	AA2005.DAT.*	APP Swedish varian
27	20	100.0	11	22	AA2006.DAT.*	Asp 1 substrate se
28	20	100.0	11	22	AA2007.DAT.*	Asp2 substrate Swe
29	20	100.0	13	13	AA2008.DAT.*	Exon XII of human
30	20	100.0	15	15	AA2009.DAT.*	HbAa-derived pept
31	20	100.0	15	22	AA2010.DAT.*	Human ribosome S7
32	20	100.0	16	21	AA2011.DAT.*	Human beta-amyloid
33	20	100.0	16	22	AA2012.DAT.*	Beta-amyloid precu
34	20	100.0	19	21	AA2013.DAT.*	Vada homologous pe
35	20	100.0	19	22	AA2014.DAT.*	Amyloid precursor
36	20	100.0	20	19	AA2015.DAT.*	Fluorogenic protea
37	20	100.0	20	20	AA2016.DAT.*	Pullanase type DBE
38	20	100.0	20	21	AA2017.DAT.*	Beta-APP alpha-sec
39	20	100.0	20	22	AA2018.DAT.*	Protease binding s
40	20	100.0	21	16	AA2019.DAT.*	Amyloid starch
41	20	100.0	21	17	AA2020.DAT.*	Potato debranching
42	20	100.0	21	18	AA2021.DAT.*	Beta-secretase sub
43	20	100.0	21	19	AA2022.DAT.*	Fluorogenic protea
44	20	100.0	21	19	AA2023.DAT.*	Fluorogenic protea
45	20	100.0	21	20	AA2024.DAT.*	Synthetic oligopep

#### ALIGNMENTS

RESULT 1	
AA1980.DAT.*	
AA1981.DAT.*	
AA1982.DAT.*	
AA1983.DAT.*	
AA1984.DAT.*	
AA1985.DAT.*	
AA1986.DAT.*	
AA1987.DAT.*	
AA1988.DAT.*	
AA1989.DAT.*	
AA1990.DAT.*	
AA1991.DAT.*	
AA1992.DAT.*	
AA1993.DAT.*	
AA1994.DAT.*	
AA1995.DAT.*	
AA1996.DAT.*	
AA1997.DAT.*	
AA1998.DAT.*	
AA1999.DAT.*	
AA2000.DAT.*	
AA2001.DAT.*	
AA2002.DAT.*	
AA2003.DAT.*	
AA2004.DAT.*	
AA2005.DAT.*	
AA2006.DAT.*	
AA2007.DAT.*	
AA2008.DAT.*	
AA2009.DAT.*	
AA2010.DAT.*	
AA2011.DAT.*	
AA2012.DAT.*	
AA2013.DAT.*	
AA2014.DAT.*	
AA2015.DAT.*	
AA2016.DAT.*	
AA2017.DAT.*	
AA2018.DAT.*	
AA2019.DAT.*	
AA2020.DAT.*	
AA2021.DAT.*	
AA2022.DAT.*	
AA2023.DAT.*	
AA2024.DAT.*	

PS Example 1; Page 7; 65pp; English.  
XX The present invention describes an antibody which recognises a  
CC neo-epitope created by the caspase mediated cleavage of amyloid-beta  
CC precursor protein (APP) or amyloid-beta precursor like protein (ABPP).  
CC These proteins are involved in the premature death of cells at sites of  
CC neurodegeneration. The antibody of the invention can thus be used in the  
CC treatment of neurodegenerative diseases such as Alzheimer's, Huntington's  
CC and Parkinson's diseases, amyotrophic lateral sclerosis, progressive  
CC multiple sclerosis, head trauma, prion related conditions,  
CC Creutzfeldt-Jacob disease, spongiform encephalopathy, Friedreich's ataxia,  
CC fatal familial insomnia, pellaeus-werzacher disease, schizophrenia,  
CC dentatorubropallidoluyian atrophy, spinocerebellar atrophy type 3,  
CC spinal bulbar muscular atrophy, spinal cord injury, stroke and brain  
CC injury.  
XX  
SQ Sequence 4 AA:  
  
Query Match 100.0%; Score 20; DB 21; Length 4;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VNLD 4  
Db 1 VNLD 4  
  
RESULT 2  
ID AAM08217 standard; peptide: 5 AA.  
XX AAM08217;  
XX  
XX AAM08217;  
XX  
XX 05-SEP-1997 (first entry)  
XX  
XX Swedish double mutant APP beta-cleavage site.  
XX  
XX Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;  
XX alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.  
XX  
XX Homo sapiens.  
XX  
XX WO9640885-A2.  
XX  
XX 19-DEC-1996.  
XX  
XX 07-JUN-1996; 96WO-US09985.  
XX  
XX 07-JUN-1995; 95US-0485152.  
XX  
XX 07-JUN-1995; 95US-0480498.  
XX  
XX (ATHE-) ATHENA NEUROSCIENCES INC.  
XX  
XX  
XX Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;  
XX Mcconlogue LC, Sinha S, Tan H;  
XX  
XX WPI: 1997-052304/05.  
XX  
XX Beta-secretase which specifically cleaves beta-amyloid precursor  
XX protein - useful to screen for inhibitors useful in treatment of  
XX Alzheimer's disease  
XX  
XX Claim 5; Page 60; 92pp; English.  
XX  
XX AAM08216, AAM08217 and AAM08350 represent beta-cleavage sites from  
XX beta-amyloid precursor proteins (APP). These sequences are recognised by  
XX the enzyme of the invention. The enzyme of the invention is  
XX beta-secretase, and specifically cleaves beta-APP at one of these sites.  
XX Normal processing of beta-APP is thought to occur via cleavage between  
XX residues 16 and 17 of the beta-amyloid peptide region by an  
XX alpha-secretase. Pathogenic processing is thought to occur by  
XX beta-secretase cleavage of beta-APP. Beta-secretase activity can be  
XX detected and measured using a method of the invention, which detects at

CC least one of the beta-secretase cleavage products formed on cleavage. The  
CC method can be used to determine whether a test substance inhibits  
CC proteolytic cleavage, by beta-secretase, of beta-APP. Compounds effective  
CC to at least partially inhibit beta-secretase activity can be used to  
CC inhibit cleavage of beta-APP in cells or mammalian hosts. Isolation and  
CC purification of beta-secretase will permit chemical modelling of a  
CC critical event in the pathology of Alzheimer's disease.  
XX  
SQ Sequence 5 AA:  
  
Query Match 100.0%; Score 20; DB 18; Length 5;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VNLD 4  
Db 1 VNLD 4  
  
RESULT 3  
ID AAM61151 standard; Peptide: 5 AA.  
XX AAM61151;  
XX  
XX AAM61151;  
XX  
XX 26-OCT-1998 (first entry)  
XX  
XX APP Swedish double mutation cleavage site.  
XX  
XX Beta-secretase; human; beta-amyloid precursor protein; APP;  
XX protease; inhibitor; screening; Alzheimer's disease; therapy.  
XX  
XX Homo sapiens.  
XX  
XX WO9826059-A1.  
XX  
XX 18-JUN-1998.  
XX  
XX 11-DEC-1996; 96WO-US19549.  
XX  
XX 11-DEC-1996; 96WO-US19549.  
XX  
XX  
XX (ATHE-) ATHENA NEUROSCIENCES INC.  
XX  
XX  
XX Anderson JP, Chrysler SMS, Keim PS, Sinha S;  
XX  
XX WPI: 1998-348519/30.  
XX  
XX Novel beta-secretase which cleaves beta-amyloid precursor protein -  
XX useful for screening for compounds which inhibit the cleavage and  
XX are useful for treating Alzheimer's disease  
XX  
XX Disclosure: Page 20; 39pp; English.  
XX  
XX This peptide comprises the site of the 'Swedish' double mutation  
XX beta-amyloid precursor protein (APP) (MPP-C125 SW) that is cleaved  
XX by a novel beta-secretase isolated from human 293 cells. This  
XX protease cleaves APP at the N-terminus of the beta-amyloid peptide  
XX (beta-AP) and is believed to be the putative beta-secretase  
XX responsible for the pathogenic processing of APP to beta-AP in  
XX Alzheimer's disease. Down's syndrome and HCHWA-D. Recombinant  
XX 125 amino acids of APP (wild-type (see AAM61150) or Swedish double  
XX mutation) fused to the C-terminal end of maltose binding protein.  
XX The fusion proteins were expressed in Escherichia coli, and used as  
XX substrates for beta-secretase in beta-secretase inhibitor assays.  
XX Compounds that inhibit APP cleavage by beta-secretase may be useful  
XX in the treatment of Alzheimer's disease.  
XX  
SQ Sequence 5 AA:  
  
Query Match 100.0%; Score 20; DB 19; Length 5;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 1 VNLD 4

RESULT 4  
 ID AAY33751  
 AAY33751 standard; Protein; 5 AA.

AC AAY33751;

DT 09-NOV-1999 (first entry)

DE Swedish mutant beta-amyloid protein precursor (APP) cleavage site.

KW Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;  
 KM Alzheimer's disease; cleavage site; mutant.

OS Homo sapiens.  
 OS Synthetic.

PN US5942400-A.

PD 24-AUG-1999.

PF 07-JUN-1996; 96US-0659984.

PR 07-JUN-1996; 96US-0659984.  
 PR 07-JUN-1995; 95US-0480498.  
 PR 07-JUN-1995; 95US-0485152.

PA (ELAN-) ELAN PHARM INC.

PI Anderson JP, Jacobson-Croak KL, Sinha S;  
 DR WPI; 1999-517417/43.

PT A method for detecting human beta-secretase cleavage of polypeptides  
 useful for identifying beta-secretase inhibitors

PS Examples; Column 28; 43pp; English.

CC This sequence is the Swedish mutant beta-amyloid protein precursor (APP)  
 CC cleavage site. APP is cleaved by beta-secretase AAY33741. The wild type  
 CC cleavage site AAY33750 and the Swedish mutant version are used in a  
 CC method for detecting human beta-secretase cleavage of polypeptides and  
 CC for identifying beta-secretase inhibitors. Inhibition of beta-secretase  
 CC activity would be useful for chemical modelling of a critical event in  
 CC the pathology of Alzheimer's disease. Inhibitors of beta-secretase would  
 CC be useful for the prevention and treatment of Alzheimer's disease and  
 CC Down's Syndrome.

SO Sequence 5 AA;

Query Match 100.0%; Score 20; DB 20; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 1 VNLD 4

RESULT 5  
 AAB47261

ID AAB47261 standard; Peptide; 5 AA.

AC AAB47261;

DT 18-JUL-2001 (first entry)

DE Swedish mutation APP sequence for cleavage by beta-secretase.

KW Beta-secretase; isotype; beta-amyloid precursor protein; APP;  
 KM beta-amyloid peptide; beta-AP; Alzheimer's disease; Down syndrome;  
 KM HCHWA-D; Swedish mutation; maltose binding protein; MBP.

OS Homo sapiens.

PN US6221645-B1.

PD 24-APR-2001.

PF 07-JUN-1996; 96US-0660531.

PR 07-JUN-1995; 95US-0480498.

PA (ELAN-) ELAN PHARM INC.

PI Chrysler SMS, Sinha S, Keim PS, Anderson JP, Tan H, McConlogue LC;  
 DR WPI; 2001-315578/33.

PT Novel antibody that specifically binds native beta-secretase protein,  
 useful for raising anti-idiotypic antibodies and for detecting or  
 diagnosing pathological conditions related to presence of respective  
 antigens

PS Example; Column 28; 42pp; English.

CC The sequences given in AAB47260-61 represent cleavage sites derived  
 CC from wild-type and the Swedish mutation of beta-amyloid precursor  
 CC protein (APP). These cleavage sites were used in fusion proteins  
 CC which were used as substrates for the beta-secretase protein which  
 CC is characterized by an ability to cleave the 695-amino acid isotype  
 CC of APP between amino acids 596 and 597. The fusion proteins contain  
 CC the carboxy-terminal end of Maltose binding protein (MBP) fused to  
 CC the carboxy-terminal 125 amino acids of either wild type APP or APP  
 CC containing the Swedish mutation. Beta-secretase is thought to be  
 CC responsible for the pathogenic processing of APP to form beta amyloid  
 CC peptide (beta-AP) in beta-AP related conditions, e.g. Alzheimer's  
 CC disease, Down syndrome, HCHWA-D etc. Beta-secretase has a molecular  
 CC weight of 260-300 kD and will bind to wheat germ agglutinin but not to  
 CC concanavalin A. Beta-secretase will cleave both the wild type and  
 CC the Swedish mutation of APP.

SO Sequence 5 AA;

Query Match 100.0%; Score 20; DB 22; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 1 VNLD 4

RESULT 6  
 ID AAY94771  
 AAY94771 standard; Protein; 8 AA.

AC AAY94771;

DT 12-FEB-2001 (first entry)

DE Beta-secretase substrate peptide SEQ ID 17.

KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease;  
 KM Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.

OS Synthetic.

PN WO200058479-A1.

```

PD 05-OCT-2000.
PF 23-MAR-2000; 2000WO-US07755.
XX
PR 26-MAR-1999; 99US-0277229.
XX
PA (AMGE-) AMGEN INC.
XX
PI Cifton M, Vassar RJ, Bennett BD;
XX
DR WPI: 2000-594643/56.
XX
XX Isolated beta-secretase nucleic acids and encoded polypeptides, useful
PT for diagnosis and gene therapy of Alzheimer's disease
XX
XX Example 10; Page 117; 145pp; English.
XX
XX This invention relates to 3 nucleotide sequences encoding beta-secretase
CC proteins. Beta-secretase is an enzyme involved in the production of one
CC of the components of amyloid plaques involved in Alzheimer's disease. The
CC invention includes an expression vector comprising the nucleotide
CC sequence, a host cell comprising the expression vector, and a process for
CC producing the protein through culturing the transformed cells. Also
CC included in the invention are a polypeptide derivative of the
CC beta-secretase protein, a fusion protein comprising beta-secretase fused
CC to a heterologous amino acid sequence, and a method for modulating the
CC levels of beta-secretase polypeptide in a mammal comprising administering
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
CC neurotropic activity. The beta-secretase nucleotide sequence may be used to
CC map locations of the beta-secretase gene and related genes on chromosomes
CC as hybridization probes in diagnostic assays to test for the presence
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease. Down's
CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
CC used as anti-sense inhibitors of beta-secretase expression, in gene
CC therapy of Alzheimer's disease, and for the identification of compounds
CC that modulate beta-secretase activity. Antibodies to the beta-secretase
CC protein may be used for in vitro and in vivo diagnostic purposes to
CC detect the presence of beta-secretase polypeptide in a body fluid or cell
CC sample. The present sequence represents a beta-secretase substrate
CC peptide.
XX
XX
SQ Sequence 8 AA:
XX
QY 1 VNLD 4
   ||||
DB 2 VNLD 5
   ||||

Query Match 100.0%; Score 20; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

RESULT 7
ID AA068473 standard; Peptide: 8 AA.
XX
AC AA068473:
XX
DT 16-JAN-2002 (first entry)
XX
DE Human Breast cancer-associated protein isoform, BPI-268 peptide #2.
XX
XX Human; Breast cancer-associated protein isoform; breast cancer;
KW immunogen; cytostatic; BPI; tryptic digest peptide.
OS Homo sapiens.
XX
XX WO200171357-A2.
PN
XX 27-SEP-2001.
PD
XX 20-MAR-2001; 2001WO-GB01219.
PF
XX

```

PR	20-MAR-2000;	2000GB-0006695.
XX	24-MAR-2000;	2000GB-0007265.
XX	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.	
PA	Herath HMC,	O'Hare MJ, Page MJ, Parekh RB, Waterfield MD:
PI	WPI: 2001-611532/70.	
XX		
DR		
PT	Identifying proteins for clinical screening, diagnosis and prognosis of	
PT	breast cancer, comprises detecting Breast Cancer-Associated Protein	
PT	Isoforms (Bpis) using two-dimensional electrophoresis -	
XX		
PS	Claim 9; Page 53; 197P: English.	
XX		
CC	The invention relates to diagnosing, determining the stage or severity,	
CC	or identifying the risk of a subject developing cancer (especially	
CC	breast cancer), or monitoring the effect of therapy on a subject with	
CC	cancer, comprising analysing a test sample using two-dimensional	
CC	electrophoresis and detecting Breast Cancer-Associated Protein	
CC	Isoforms (Bpis). The methods disclosed are used for the diagnosis and	
CC	prognosis of breast cancer, for determining the severity of breast	
CC	cancer, and for identifying the effect of therapy administered to a subject,	
CC	Antibodies raised against the binding domain of a BPI, the binding domain	
CC	of a BPI, a nucleic acid encoding a BPI, or a nucleic acid that inhibits	
CC	the function of a BPI can be incorporated into a pharmaceutical	
CC	composition for treating or preventing breast cancer. The methods use	
CC	sensitive and specific biomarkers provide early diagnosis of breast	
CC	cancer, and the compositions are more potent, specific, and has a more	
CC	rapid effect with fewer side effects than other prior art methods.	
CC	The present sequence is a tryptic digest peptide from a BPI of the	
CC	invention.	
XX		
SQ	Sequence	8 AA:
QY	Query Match	100.0%; Score 20; DB 22; Length 8;
	Best Local Similarity	100.0%; Pred. No. 6.4e+05;
Matches	4; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
DY	1 VNLD 4	
DB	3 VNLD 6	
RESULT 8		
AAEI0661	AAEI0661 standard; peptide; 8 AA.	
ID		
AC	AAEI0661;	
XX		
DT	10-DEC-2001 (first entry)	
XX		
DE	Human aspartyl protease-1 beta-secretase Swedish mutant peptide.	
XX		
XX	Human; aspartyl protease 1; Aspl; amyloid precursor protein; APP;	
KW	Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;	
KW	amyloid plaque; neuronal loss; proteolytic; nootropic; neuroprotective;	
KW	aspartyl protease-1 beta-secretase Swedish mutant peptide.	
XX		
OS	Homo sapiens.	
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FM	Cleavage-site	4..5
XX		
PN	GB2357767-A.	
XX		
PD	04-JUL-2001.	
XX		
PF	22-SEP-2000; 2000GB-0023315.	
XX		
XX	23-SEP-1999; 99US-0155493.	
PR		

PR 23-SEP-1999; 99US-0404133.  
PR 23-SEP-1999; 99WO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Blenkowski MJ, Gurney M;  
XX  
DR WPI; 2001-444208/48.  
XX  
PT Polypeptide comprising fragments of human aspartyl protease with  
PT amyloid precursor protein processing activity and alpha-secretase  
PT activity, for identifying modulators useful in treating Alzheimer's  
PT disease -  
XX  
PS Example 15; Page 92; 187pp; English.  
XX  
CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified  
CC Asp1 proteins which lack transmembrane domain or amino terminal  
CC domain or cytoplasmic domain and retains alpha-secretase activity  
CC and amyloid protein precursor (APP) processing activity. The proteins  
CC of the invention are useful for assaying hu-Asp1 alpha-secretase  
CC activity, which in turn is useful for identifying modulators of  
CC hu-Asp1 alpha-secretase activity, where modulators that increase  
CC hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's  
CC disease (AD) which causes progressive dementia with consequent  
CC formation of amyloid plaques, neurofibrillary tangles, gliosis and  
CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying  
CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with  
CC the substrate under acidic conditions and determining the level of  
CC hu-Asp1 proteolytic activity. The present sequence is human aspartyl  
CC protease-1 (hu-Asp-1) beta-secretase Swedish (Sw) mutant peptide  
CC which is used for determining the enzymatic activity of Asp-1 protein  
CC lacking a transmembrane (TM) domain and containing (His)<sub>6</sub> tag.  
XX  
SQ Sequence 8 AA:  
  
Query Match 100.0%; Score 20; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VNLD 4  
DB 2 VNLD 5  
  
RESULT 9  
AAE02613  
ID AAE02613 standard; peptide; 8 AA.  
XX  
AC AAE02613;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Human Aspartyl protease-1 beta-secretase Swedish mutant form peptide.  
XX  
KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;  
KW Alzheimer's disease; antialzheimer's; aspartyl protease 1; Asp1;  
KW beta-secretase.  
XX  
OS Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
XX FT Cleavage-site 4..5  
XX  
XX PN WO200123533-A2.  
XX  
XX PD 05-APR-2001.  
XX  
XX PF 22-SEP-2000; 2000WO-US26080.  
XX  
XX PR 23-SEP-1999; 99US-0155493.  
XX  
CC

PR 23-SEP-1999; 99WO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Gurney M, Blenkowski MJ;  
XX  
DR WPI; 2001-290516/30.  
XX  
XX  
PT Enzymes that cleave the alpha-secretase site of the amyloid precursor  
PT protein, useful for the treatment of Alzheimer's disease -  
XX  
PS Example 15; Page 94; 189pp; English.  
XX  
CC The present invention relates to enzymes for cleaving the alpha-  
CC secretase site of the amyloid precursor protein (APP) and methods of  
CC identifying those enzymes. The methods may be used to identify enzymes  
CC that may be used to cleave the alpha-secretase cleavage site of the APP  
CC protein. The enzymes may be used to treat or modulate the progress of  
CC Alzheimer's disease. The present sequence is human aspartyl protease-1  
CC (hu-Asp-1) beta-secretase, Swedish (Sw) mutant form peptide which is used  
CC for determining the enzymatic activity of Asp-1 delatm (His)<sub>6</sub> protein.  
XX  
SQ Sequence 8 AA:  
  
Query Match 100.0%; Score 20; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VNLD 4  
DB 2 VNLD 5  
  
RESULT 10  
AAW82085  
ID AAW82085 standard; peptide; 9 AA.  
XX  
AC AAW82085;  
XX  
DT 18-FEB-1999 (first entry)  
XX  
DE Fluorogenic protease indicator protease binding peptide #63.  
XX  
KW Protease activity; fluorophore; detection; fluorogenic; cellular uptake;  
KW conformation change.  
XX  
OS Synthetic.  
XX  
XX WO9837226-A1.  
XX  
XX PN 27-AUG-1998.  
XX  
PD 20-FEB-1998; 98WO-US03000.  
XX  
PF 20-FEB-1997; 97US-0802981.  
XX  
PR (ONCO-) ONCOIMMUNIN INC.  
XX  
XX PA Komoriya A, Packard BS;  
XX  
XX PI  
XX  
XX DR WPI; 1998-467579/40.  
XX  
XX  
XX PT New fluorogenic compositions - containing 2 fluorophores separated  
XX by a peptide comprising a protease binding site, used for detecting  
XX protease activity in samples.  
XX  
PS Claim 4; Page 77; 90pp; English.  
XX  
XX AAW82023-W82240 are peptides used in the construction of a fluorogenic  
XX CC composition which is used for the detection of protease activity in  
XX CC biological samples. The products can be used for the detection of

conformation changes in nucleic acids, oligosaccharides, polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids, glycoproteins, steroids or polymers. In addition, attachment of a hydrophobic group to a molecule can be used to enhance uptake by cells. The composition is composed of P = peptide comprising a protease binding site for the protease, F1, F2 peptides = fluorophores where F1 is attached to the amino terminal amino acid and F2 is attached to the carboxyl terminal amino acid and S1, S2 peptides = when present, are attached to the amino terminal amino acid and S1, S2 peptides = when present, is attached to the amino terminal amino acid, and S2, when present, is attached to the carboxyl terminal amino acid.

Sequence 9 AA;

Query Match 100.0%; Score 20; DB 19; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
1111  
DB 3 VNLD 6

RESULT 11

AAW82081  
ID AAW82081 standard; peptide; 9 AA.

XX AAW82081;

XX 18-FEB-1999 (first entry)

XX Fluorogenic protease indicator protease binding peptide #59.

XX Protease activity; fluorophore; detection; fluorogenic; cellular uptake; conformation change.

XX Synthetic.

XX WO9837226-A1.

XX 27-AUG-1998.

XX 20-FEB-1998; 98WO-US03000.

XX 20-FEB-1997; 97US-0802981.

XX (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI: 1998-467579/40.

XX New fluorogenic compositions - containing 2 fluorophores separated by a peptide comprising a protease binding site, used for detecting protease activity in samples.

XX Claim 4; Page 77; 90pp; English.

XX AAW82023-W82240 are peptides used in the construction of a fluorogenic composition which is used for the detection of protease activity in biological samples. The products can be used for the detection of conformation changes in nucleic acids, oligosaccharides, polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids, glycoproteins, steroids or polymers. In addition, attachment of a hydrophobic group to a molecule can be used to enhance uptake by cells. The composition is composed of P = peptide comprising a protease binding site for the protease, F1, F2 peptides = fluorophores where F1 is attached to the amino terminal amino acid and F2 is attached to the carboxyl terminal amino acid and S1, S2 peptides = when present, are attached to the amino terminal amino acid and S1, S2 peptides = when present, is attached to the amino terminal amino acid, and S2, when present, is attached to the carboxyl terminal amino acid.

SQ Sequence 9 AA;

Query Match 100.0%; Score 20; DB 19; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
1111  
DB 3 VNLD 6

RESULT 12

AAW82081  
ID AAW82081 standard; peptide; 9 AA.

XX AAW82081;

XX 14-NOV-2000 (first entry)

XX A peptide fragment derived from beta-amyloid precursor protein.

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide; amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

XX inhibitor.

XX Homo sapiens.

XX WO200047618-A2.

XX 17-AUG-2000.

XX 10-FEB-2000; 2000WO-US03819.

XX 10-FEB-1999; 99US-0119571.

XX 15-JUN-1999; 99US-0139172.

XX (ELAN-) ELAN PHARM INC.

XX Anderson JP, Bast G, Doane MT, Frigon N, John V, Power M;

XX Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX WPI: 2000-533011/48.

XX Purified beta-secretase protein used in assays to discover inhibitors which can be used for the treatment of amyloidogenic diseases e.g. Alzheimer's disease -

XX Disclosure; Page 12; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves beta-amyloid precursor protein to produce beta-amyloid peptide. This enzyme is therefore implicated in the production of amyloid plaque components which accumulate in the brains of individuals afflicted with CC Alzheimer's disease. Inhibitors of beta-secretase are administered to CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-like pathology to test if they maintain or improve cognitive CC ability or reduce the plaque burden. The compounds are used for the CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The CC present sequence represents a peptide derived from beta-amyloid CC precursor protein

XX Sequence 9 AA;

Query Match 100.0%; Score 20; DB 21; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
1111  
DB 3 VNLD 6

RESULT 13

AAB07894  
ID AAB07894 standard; Peptide; 9 AA.  
XX  
AC AAB07894;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE Substrate for beta-secretase enzyme.  
XX  
KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KM amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
XX Inhibitor.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Cleavage-site 5..6  
XX  
PN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
XX  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
XX  
PS Example 4: Page 71; 121pp; English.  
XX  
CC The specificity of the beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide substrate used to test the  
CC activity of beta-secretase enzyme.  
XX  
SQ Sequence 9 AA;  
XX  
Query Match 100.0%; Score 20; DB 21; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 VNLD 4  
XX  
DB 3 VNLD 6  
XX  
RESULT 14  
AAG73297  
ID AAG73297 standard; Peptide; 9 AA.  
XX  
AC AAG73297;  
XX  
DT 14-AUG-2001 (first entry)  
XX  
DE Protease indicator compound peptide #26.  
XX

KM Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KM viral infection; cancer metastasis; emphysema; arthritis;  
KM thrombosis; haemophilia.  
XX  
XX  
OS Synthetic.  
XX  
PN WO200118238-A1.  
XX  
PD 15-MAR-2001.  
XX  
PF 11-SEP-2000; 2000WO-US24882.  
XX  
PR 10-SEP-1999; 99US-0394019.  
XX  
XX (ONCO-) ONCOMUNIN INC.  
XX  
PI Komoriya A, Packard BS;  
XX  
PI WPI: 2001-389573/41.  
XX  
DR New fluorogenic compositions whose fluorescence level increases in the  
XX presence of active proteases, useful for detecting and localizing  
XX protease activity in biological samples, particularly in frozen tissue  
XX samples -  
XX  
PS Claim 1: Page 70; 86pp; English.  
XX  
XX  
CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
XX exemplification of the invention.  
XX  
SQ Sequence 9 AA;  
XX  
Query Match 100.0%; Score 20; DB 22; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 VNLD 4  
XX  
DB 3 VNLD 6  
XX  
RESULT 15  
AAW08362  
ID AAW08362 standard; peptide; 10 AA.  
XX  
AC AAW08362;  
XX  
DT 05-SEP-1997 (first entry)  
XX  
DE Beta-secretase substrate #3.  
XX  
KM Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;  
KM alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1 /note="acetylated"  
XX  
XX  
PN WO9640885-A2.  
XX  
PD 19-DEC-1996.  
XX  
PF 07-JUN-1996; 96WO-US09985.  
XX  
PF 07-JUN-1995; 95US-0485152.  
XX  
PR

PR 07-JUN-1995; 95US-0480498.

XX (ATHE-) ATHENA NEUROSCIENCES INC.

XX Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;

PI Mcconlogue JC, Sinha S, Tan H;

XX WPI; 1997-052304/05.

DR Beta-secretase which specifically cleaves beta-amyloid precursor

XX protein - useful to screen for inhibitors useful in treatment of

PT Alzheimer's disease

PS Disclosure; Page 45; 92pp; English.

XX AAM08359-W08362 represent substrates for the enzyme of the invention.

CC The enzyme of the invention is beta-secretase, and specifically cleaves

CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP

CC is thought to occur via cleavage between residues 16 and 17 of the

CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing

CC is thought to occur by beta-secretase cleavage of beta-APP.

CC Beta-secretase activity can be detected and measured using a method of

CC the invention, which detects at least one of the beta-secretase cleavage

CC products formed on cleavage. The method can be used to determine whether

CC a test substance inhibits proteolytic cleavage, by beta-secretase, of

CC beta-APP. Compounds effective to at least partially inhibit

CC beta-secretase activity can be used to inhibit cleavage of beta-APP in

CC cells or mammalian hosts. Isolation and purification of beta-secretase

CC will permit chemical modelling of a critical event in the pathology of

CC Alzheimer's disease.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Search completed: October 30, 2002, 12:27:11  
Job time : 2.60197 secs

SO Sequence 10 AA;  
Query Match 100.0%; Score 20; DB 18; Length 10;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VNLD 4  
| | | |  
Db 3 VNLD 6



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 ; Search time 0.570025 Seconds  
(without alignments)  
171,400 Million cell updates/sec

Title: US-09-724-571-104

Perfect score: 20  
Sequence: 1 VNLD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	4	US-09-054-334-6	Sequence 6, Appli
2	20	100.0	5	US-08-480-498-2	Sequence 2, Appli
3	20	100.0	5	US-08-659-984A-14	Sequence 14, Appli
4	20	100.0	5	US-08-660-531-14	Sequence 14, Appli
5	20	100.0	5	US-09-054-334-2	Sequence 2, Appli
6	20	100.0	9	US-08-802-981-219	Sequence 219, App
7	20	100.0	9	US-08-802-981-223	Sequence 223, App
8	20	100.0	10	US-08-659-984A-19	Sequence 19, Appli
9	20	100.0	10	US-08-660-531-19	Sequence 19, Appli
10	20	100.0	11	PCT-US94-07043A-3	Sequence 19, Appli
11	20	100.0	20	US-09-196-293-2	Sequence 2, Appli
12	20	100.0	20	US-08-209-603E-2	Sequence 2, Appli
13	20	100.0	21	US-08-659-984A-18	Sequence 18, Appli
14	20	100.0	21	US-08-596-257A-1	Sequence 18, Appli
15	20	100.0	21	US-08-802-981-112	Sequence 11, Appli
16	20	100.0	21	US-08-802-981-116	Sequence 11, App
17	20	100.0	21	US-08-860-339-1	Sequence 11, App
18	20	100.0	21	US-08-650-531-18	Sequence 18, Appli
19	20	100.0	30	US-08-659-984A-17	Sequence 17, Appli
20	20	100.0	30	US-08-660-531-17	Sequence 17, Appli
21	20	100.0	33	US-08-659-984A-16	Sequence 16, Appli
22	20	100.0	33	US-08-660-531-16	Sequence 16, Appli
23	20	100.0	42	US-08-659-984A-15	Sequence 15, Appli
24	20	100.0	42	US-08-660-531-15	Sequence 15, Appli
25	20	100.0	67	US-08-588-258B-6	Sequence 6, Appli
26	20	100.0	67	US-08-460-505-6	Sequence 6, Appli
27	20	100.0	67	PCT-US96-08295-6	Sequence 6, Appli

28	20	100.0	68	2	US-08-588-258B-10	Sequence 10, Appli
29	20	100.0	68	3	US-08-460-505-10	Sequence 10, Appli
30	20	100.0	68	5	PCT-US96-08295-10	Sequence 10, Appli
31	20	100.0	87	1	US-08-204-740-13	Sequence 13, Appli
32	20	100.0	87	3	US-09-081-167A-13	Sequence 13, Appli
33	20	100.0	87	3	US-09-081-395-13	Sequence 13, Appli
34	20	100.0	87	4	US-09-416-833-13	Sequence 13, Appli
35	20	100.0	87	5	PCT-US95-02521-13	Sequence 13, Appli
36	20	100.0	103	3	US-08-339-708A-12	Sequence 12, Appli
37	20	100.0	108	3	US-08-204-740-11	Sequence 11, Appli
38	20	100.0	108	3	US-09-081-167A-11	Sequence 11, Appli
39	20	100.0	108	3	US-09-081-395-11	Sequence 11, Appli
40	20	100.0	108	4	US-09-416-833-11	Sequence 11, Appli
41	20	100.0	108	5	PCT-US95-02521-11	Sequence 11, Appli
42	20	100.0	120	4	US-08-890-865A-13	Sequence 13, Appli
43	20	100.0	120	4	US-08-890-865A-14	Sequence 14, Appli
44	20	100.0	146	4	US-08-858-207A-494	Sequence 494, App
45	20	100.0	159	3	US-09-010-809-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-054-334-6  
; Sequence 6, Application US/09054334  
; Patent No. 6329163  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Jacobson-Croak, Kirsten L.  
; APPLICANT: Sinha, Sukanto  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
; TITLE OF INVENTION: Inhibition  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/054,334  
; FILING DATE: 02-APR-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,152  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 015270-002820US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-054-334-6

Query Match 100.0%; Score 20; DB 4; Length 4;  
Best Local Similarity 100.0%; Prod. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
1111  
Db 1 VNLD 4

RESULT 2  
US-08-480-498-2

Sequence 2, Application US/08480498  
Patent No. 5744346  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: One Market Plaza, Stewart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,498  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-002200  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-480-498-2

Query Match 100.0%; Score 20; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
1111  
Db 1 VNLD 4

RESULT 3  
US-08-659-984A-14

Sequence 14, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California

COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A

FILING DATE: 07-JUN-1996

CLASSIFICATION: 436

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.

REGISTRATION NUMBER: 29,541

REFERENCE/DOCKET NUMBER: 15270-002810US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400

TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-14

Query Match 100.0%; Score 20; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
1111  
Db 1 VNLD 4

RESULT 4  
US-08-660-531-14

Sequence 14, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-14

Query Match  
Best Local Similarity 100.0%; Score 20; DB 4; Length 5;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
||||  
DB 1 VNLD 4

RESULT 5  
US-09-054-334-2  
Sequence 2, Application US/09054334  
Patent No. 6329163  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,334  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-002820US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-054-334-2

Query Match  
Best Local Similarity 100.0%; Score 20; DB 4; Length 5;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VNLD 4  
||||

DB 1 VNLD 4

RESULT 6  
US-08-802-981-219  
Sequence 219, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 219:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-802-981-219

Query Match  
Best Local Similarity 100.0%; Score 20; DB 3; Length 9;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
||||  
DB 3 VNLD 6

RESULT 7  
US-08-802-981-223  
Sequence 223, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 223:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-802-981-223

Query Match  
Best Local Similarity 100.0%; Score 20; DB 3; Length 9;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 UNLD 4  
1111  
DB 3 UNLD 6

RESULT 8  
US-08-659-984A-19  
Sequence 19, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2422  
TELEFAX: 415-326-2400  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-659-984A-19

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 10;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 UNLD 4  
1111  
DB 3 UNLD 6

RESULT 9  
US-08-660-531-19  
Sequence 19, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-660-531-19

Query Match  
Best Local Similarity 100.0%; Score 20; DB 4; Length 10;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 UNLD 4

||||  
DB 3 VMLD 6

## RESULT 10

PCT-US94-07043A-3  
Sequence 3, Application PC/TUS9407043A

## GENERAL INFORMATION:

APPLICANT: Tamburini, Paul P.; Benz, G nter; H bich,  
APPLICANT: Dieler, Robert N.; Koenig, Gerhard  
TITLE OF INVENTION: CATHEPSIN D IS AN AMYLOIDOTIC  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Miles Inc.  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06516

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: Sharp PC 4600  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07043A  
FILING DATE: June 21, 1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10889  
FILING DATE: November 12, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/995,660  
FILING DATE: December 16, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/880,914  
FILING DATE: May 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pamela A. Slomton  
REGISTRATION NUMBER: 31,060  
REFERENCE/DOCKET NUMBER: MTI 224.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 937-2340  
TELEFAX: (203) 937-2795  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US94-07043A-3

Query Match 100.0%; Score 20; DB 5; Length 11;  
Best Local Similarity 100.0%; Pred. No. 26;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VMLD 4  
||||  
DB 4 VMLD 7

## RESULT 11

US-09-196-293-2  
Sequence 2, Application US/09196293

## GENERAL INFORMATION:

APPLICANT: Fuchs, Renate  
APPLICANT: Motz, Manfred  
APPLICANT: Soutscheck, Erwin  
APPLICANT: Wilske, Bettina  
APPLICANT: Preac-Mursic, Vera  
TITLE OF INVENTION: Active proteins from Borrelia  
TITLE OF INVENTION: burgdorferi

## FILE REFERENCE: 738 001US2

CURRENT APPLICATION NUMBER: US/09/196,293  
CURRENT FILING DATE: 1998-11-19  
EARLIER APPLICATION NUMBER: US 08/209,603  
EARLIER FILING DATE: 1994-03-10  
EARLIER APPLICATION NUMBER: US 07/862,535  
EARLIER FILING DATE: 1992-06-19  
EARLIER APPLICATION NUMBER: WO PCT/EP90/02282  
EARLIER FILING DATE: 1990-12-21  
EARLIER APPLICATION NUMBER: DE P39 42 728.5  
EARLIER FILING DATE: 1989-12-22  
EARLIER APPLICATION NUMBER: DE P40 18 988.0  
EARLIER FILING DATE: 1990-06-13  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Borrelia burgdorferi  
US-09-196-293-2

## Query Match

Best Local Similarity 100.0%; Score 20; DB 4; Length 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VMLD 4  
||||  
DB 11 VMLD 14

## RESULT 12

US-08-209-603E-2  
Sequence 2, Application US/08209603E

## GENERAL INFORMATION:

APPLICANT: FUCHS, RENATE  
APPLICANT: WILSKE, BETTINA  
APPLICANT: PREAC-MURISIC, VERA  
APPLICANT: MOTZ, MANFRED  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PROTEINS  
TITLE OF INVENTION: FROM BORRELLIA BURGDORFERI  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/209,603E  
FILING DATE: 10-MAR-1994  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP90/02282  
FILING DATE: 21-DEC-1990  
APPLICATION NUMBER: US 07/862,535  
FILING DATE: 19-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9217-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 20  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
DESCRIPTION: N/A  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: N/A  
ORIGINAL SOURCE: B. BURGDOFFER  
IMMEDIATE SOURCE: LIBRARY: DSM 5662  
POSITION IN GENOME: N/A  
FEATURE:  
IDENTIFICATION METHOD: amino acid analysis  
PUBLICATION INFORMATION: N/A  
US-08-209-603E-2

Query Match 100.0%; Score 20; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
DB 11 VNLD 14

RESULT 13  
US-08-659-984A-18  
; Sequence 18, Application US/08655984A  
; Patent No. 5942400  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Jacobson-Croak, Kirsten L.  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
; TITLE OF INVENTION: Inhibition  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,984A  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,152  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002810US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-659-984A-18

Query Match 100.0%; Score 20; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
DB 3 VNLD 6

RESULT 14  
US-08-596-257A-1  
; Sequence 1, Application US/08596257A  
; Patent No. 6001628  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMAN, Jens  
; APPLICANT: EMMERMAN, Michael  
; APPLICANT: VIRGIN, Ivai  
; TITLE OF INVENTION: BRANCHING ENZYMES AND DNA SEQUENCES  
; TITLE OF INVENTION: CODING THEM, SUITABLE FOR CHANGING THE DEGREE OF BRANCHING  
; TITLE OF INVENTION: OF AMYLOPECTIN STARCH IN PLANTS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/596,257A  
; FILING DATE: 08-FEB-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 4327165.0  
; FILING DATE: 09-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meilman, Edward A.  
; REGISTRATION NUMBER: 24,735  
; REFERENCE/DOCKET NUMBER: P/951-118  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 382-0700  
; TELEFAX: (212) 382-0888  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-596-257A-1

Query Match 100.0%; Score 20; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
DB 5 VNLD 8

RESULT 15  
US-08-802-981-112  
; Sequence 112, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.

TITLE OF INVENTION: Compositions for the Detection of Enzyme  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Alb"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-112

Query Match 100.0%; Score 20; DB 3; Length 21;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VNLD 4  
|||  
Db 8 VNLD 11

Search completed: October 30, 2002, 12:32:37  
Job time : 1.57002 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 0.766585 Seconds

(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-104

Perfect score: 20

Sequence: 1 VNLD 4

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	20	100.0	15	2	P00017
2	20	100.0	30	2	S27306
3	20	100.0	33	2	H95094
4	20	100.0	33	2	E82852
5	20	100.0	49	2	T07309
6	20	100.0	59	2	T38172
7	20	100.0	62	2	S45295
8	20	100.0	62	2	S27308
9	20	100.0	63	2	S04973
10	20	100.0	72	2	S34404
11	20	100.0	79	2	T17694
12	20	100.0	80	2	H89990
13	20	100.0	83	2	T36483
14	20	100.0	84	2	E70841
15	20	100.0	88	2	S50644
16	20	100.0	89	2	J54814
17	20	100.0	89	2	H87367
18	20	100.0	90	2	T46007
19	20	100.0	91	2	S57268
20	20	100.0	91	2	T29686
21	20	100.0	91	2	A75188
22	20	100.0	91	2	A75188
23	20	100.0	93	2	G81410
24	20	100.0	96	2	PC4086
25	20	100.0	97	2	S72857
26	20	100.0	99	2	T06983
27	20	100.0	104	2	A69407
28	20	100.0	104	2	E86263
29	20	100.0	105	2	T42021

#### ALIGNMENTS

30	20	100.0	106	2	A10113	probable Rieseke pr
31	20	100.0	106	2	JG0021	flagellar hook-bas
32	20	100.0	109	2	B70637	hypothetical prote
33	20	100.0	110	2	S65003	hypothetical prote
34	20	100.0	112	2	A69451	conserved hypotet
35	20	100.0	112	2	B95061	BLPS protein limpo
36	20	100.0	112	2	P97929	regulatory protein
37	20	100.0	115	2	T48186	light-inducible pr
38	20	100.0	116	1	NGNXXI	nerve growth facto
39	20	100.0	116	2	A58566	N-acetylmuramoyl-L
40	20	100.0	116	2	G64126	conserved hypotet
41	20	100.0	117	2	F64708	holo-(acyl carrier
42	20	100.0	119	2	F83714	probable sensory t
43	20	100.0	120	2	B95269	ribosomal protein
44	20	100.0	121	2	S78265	hypothetical prote
45	20	100.0	123	2	C81004	

RESULT 1  
P00017  
terminal protein - phage M2 (fragment)  
C:Species: phage M2  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Sep-1999  
C:Accession: P00017  
R:Matsumoto, K.; Takano, H.; Kim, C.I.; Hirokawa, H.  
Gene 84, 247-255, 1989  
A:Title: Primary structure of bacteriophage M2 DNA polymerase: conserved segments wit  
A:Reference number: J00161; MUID:90128268  
A:Accession: P00017  
A:Molecule type: DNA  
A:Residues: 1-15 <MAT>  
A:Cross-references: GB:M33144; NID:g215507; PIDN:AAA3367.1; PID:g215508  
C:Genetics:  
A:Gene: E  
C:Superfamily: phage PZA terminal protein

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 15;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
Db 8 VNLD 11

#### RESULT 2

S27306  
surface layer protein - Aeromonas hydrophila  
N:Alternate names: S-layer protein  
C:Species: Aeromonas hydrophila  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 07-May-1999  
R:Dooley, J.S.G.; McCubbin, W.D.; Kay, C.M.; Trust, T.J.  
J. Bacteriol. 170, 2631-2638, 1988  
A:Title: Isolation and biochemical characterization of the S-layer protein from a pat  
A:Reference number: S27306; MUID:88227842  
A:Accession: S27306  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-30 <DOO>

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 30;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
Db 1 VNLD 4

RESULT 3  
H95094  
hypothetical protein SP0821 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: H95094  
R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held  
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
non, T.; Hickey, E.K.; Holt, I.E.  
Science 293: 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: H95094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-33 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK74953.1; PID:914972294; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0821

Query Match 100.0%; Score 20; DB 2; Length 33;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
| | | |  
Db 5 VNLD 8

RESULT 4  
E82852  
hypothetical protein XF0065 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82852  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406: 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82852  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-33 <STM>  
A:Cross-references: GB:AE003860; GB:AE003849; NID:99104830; PIDN:AAF82878.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H  
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to Genbank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
Rodrigues, V.; Rosa, A.O. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF0065

Query Match 100.0%; Score 20; DB 2; Length 33;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
| | | |  
Db 15 VNLD 18

RESULT 5  
T07309  
hypothetical protein 49e - Chlorella vulgaris chloroplast  
C:Species: Chlorella vulgaris  
C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000  
C:Accession: T07309  
R:Wakagagi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Na  
Proc. Natl. Acad. Sci. U.S.A. 94: 5967-5972, 1997  
A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga C  
Proc. Natl. Acad. Sci. U.S.A. 94: 5967-5972, 1997  
A:Reference number: Z15985; MUID:97303241  
A:Accession: T07309  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-49 <NAK>  
A:Cross-references: EMBL:AB001684; NID:92224352; PIDN:BA57957.1; PID:92224473  
C:Genetics:  
A:Genome: chloroplast  
C:Keywords: chloroplast

Query Match 100.0%; Score 20; DB 2; Length 49;  
Best Local Similarity 100.0%; Pred. No. 1,3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
| | | |  
Db 3 VNLD 6

RESULT 6  
T38172  
hypothetical protein SPAC22E12.15 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T38172  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z21775  
A:Accession: T38172  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-59 <DEV>  
A:Cross-references: EMBL:Z70043; NID:91220275; PIDN:CAA93902.1; GSPDB:GN00066; SPDB:S  
A:Experimental source: strain 972h; cosmid c22E12  
C:Genetics:  
A:Gene: SPDB:SPAC22E12.15  
A:Map position: 1  
A:introns: 19/1: 23/1

Query Match 100.0%; Score 20; DB 2; Length 59;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
| | | |  
Db 19 VNLD 22

RESULT 7  
S45295  
cyclin Cyc3 - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 13-Jan-1995 #sequence\_revision 19-Apr-1996 #text\_change 16-Jul-1999  
C:Accession: S45295  
R:Day, I.S.; Reddy, A.S.N.  
Biochim. Biophys. Acta 1218: 115-118, 1994  
A:Title: Cloning of a family of cyclins from Arabidopsis thaliana.  
A:Reference number: S45293; MUID:94250688  
A:Accession: S45295  
A:Molecule type: mRNA  
A:Residues: 1-62 <DAY>  
A:Cross-references: EMBL:L27225; NID:9456021; PIDN:AAA19880.1; PID:9456022  
C:Superfamily: cyclin

C:Keywords: cell cycle control

Query Match 100.0%; Score 20; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
||||  
DB 27 VNLD 30

RESULT 8  
S27308  
Surface layer protein - Aeromonas hydrophila

N:Alternate names: S-layer protein  
C:Species: Aeromonas hydrophila  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-1999  
C:Accession: S27308  
R:Kosorzynska, M.; Dooley, J.S.G.; Shimojo, T.; Sakata, T.; Trust, T.J.  
J. Bacteriol. 174, 40-47, 1992

A:Title: Antigenic diversity of the S-layer proteins from pathogenic strains of Aeromonas  
A:Reference number: S27308; MUID:92104988  
C:Accession: S27308

A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-62 <MOS>

Query Match 100.0%; Score 20; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
||||  
DB 1 VNLD 4

RESULT 9  
S04973

Antifreeze protein AB1 - Antarctic eelpout

C:Species: Austrolyctichys brachycephalus (Antarctic eelpout)  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 24-Oct-2000  
C:Accession: S04973; B30095

R:Cheng, C.H.C.; Devries, A.L.  
Biochim. Biophys. Acta 997, 55-64, 1989

A:Title: Structures of antifreeze peptides from the antarctic eel pout, Austrolyctichys

A:Reference number: S04973; MUID:893323219  
C:Accession: S04973

A:Molecule type: protein  
A:Residues: 1-63 <CHE>

A>Note: the accession number A30095 cited as assigned by EMBL is not in GenBank release  
C:Superfamily: antifreeze protein SPI

C:Keywords: antifreeze

Query Match 100.0%; Score 20; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
||||  
DB 48 VNLD 51

RESULT 10  
S34404

ribosomal protein L15 - Bacillus licheniformis (fragment)

C:Species: Bacillus licheniformis

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Mar-1999  
C:Accession: S34404

R:Tschauder, S.; Driessen, A.J.M.; Freudl, R.  
Mol. Gen. Genet. 235, 147-152, 1992

A:Title: Cloning and molecular characterization of the secY genes from Bacillus lichenif  
A:Reference number: S30115; MUID:93062802  
C:Accession: S34404

A:Molecule type: DNA  
A:Residues: 1-72 <TSC>  
A:Cross-references: EMBL:X70087  
C:Genetics:  
A:Gene: rpl15  
C:Superfamily: Escherichia coli ribosomal protein L15  
C:Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 20; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
||||  
DB 3 VNLD 6

RESULT 11  
T17694

hypothetical protein a204L - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17694

R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999

A:Reference number: T17694  
C:Accession: T17694

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-79 <GRA>  
A:Cross-references: EMBL:U42580; NID:94028896; PIDN:AMC96572.1

A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:

A>Note: a204L

Query Match 100.0%; Score 20; DB 2; Length 79;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
||||  
DB 48 VNLD 51

RESULT 12  
H89990

hypothetical protein SA1809 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: H89990

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani-Tl, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
C. Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
C:Accession: H89990

A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-80 <KUR>  
A:Cross-references: GB:BA000018; PID:G13701796; PIDN:BA843089.1; GSPDB:GN00149

A:Experimental source: strain N315  
C:Genetics:

A:Gene: SA1809

Query Match 100.0%; Score 20; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
||||  
DB 67 VNLD 70

RESULT 13  
T36483

probable membrane protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: J36483  
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1999

A:Reference number: 221608  
A:Accession: T36483

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-83 <SAD>

A:Cross-references: EMBL:AL096622; PIDN:CAB46925.1; GSPDB:GN00070; SCOEDB:SCGD3.03

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCGD3.03

Query Match 100.0%; Score 20; DB 2; Length 83;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
Db 27 VNLD 30

RESULT 14  
E70841

hypothetical protein RV2104c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: E70841

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: E70841

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-84 <COL>

A:Cross-references: GB:AL021924; GB:AL123456; NID:g3261519; PIDN:CA17287.1; PID:el25238

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2104c

Query Match 100.0%; Score 20; DB 2; Length 84;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
Db 59 VNLD 62

RESULT 15  
JS0644

eclosion hormone precursor - silkworm

C:Species: Bombyx mori (silkworm)

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 21-Jul-2000

C:Accession: JS0644

R:Kimoto, T.; Tanaka, H.; Sato, B.; Nagasawa, H.; Suzuki, A.

Biochem. Biophys. Res. Commun. 182, 514-519, 1992

A:Title: Nucleotide sequence of cDNA for the eclosion hormone of the silkworm, Bombyx mori

A:Reference number: JS0644; MUID:92134263

A:Accession: JS0644

A:Molecule type: mRNA

A:Residues: 1-88 <KAM>

A:Cross-references: GB:D10135; NID:g217271; PIDN:BAA01012.1; PID:g217272

C:Comment: This neuropeptide controls ecdysis.  
C:Superfamily: eclosion hormone  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-88/Product: eclosion hormone #status predicted <ECL>

Query Match 100.0%; Score 20; DB 2; Length 88;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
Db 19 VNLD 22

Search completed: October 30, 2002, 12:31:41  
Job time : 3.7658 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:10 ; Search time 0.34398 Seconds

(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-104

Perfect score: 20

Sequence: 1 VNLD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	15	1	TERM_BPM2
2	20	100.0	40	1	CSP4_STRMT
3	20	100.0	59	1	YDBF_SCHRO
4	20	100.0	63	1	ANP1_AUSBR
5	20	100.0	67	1	RGS3_RAT
6	20	100.0	72	1	RL15_BACLI
7	20	100.0	88	1	ECIL_BOMMO
8	20	100.0	90	1	EP1B_SURSO
9	20	100.0	91	1	EP1B_PYRAB
10	20	100.0	96	1	RL15_STRSC
11	20	100.0	104	1	SR19_ARCFU
12	20	100.0	105	1	HAL2_CANAL
13	20	100.0	106	1	FLIE_BACSU
14	20	100.0	112	1	SPEH_ARCFU
15	20	100.0	116	1	NGF_NAJAT
16	20	100.0	116	1	NGF_NAJAT
17	20	100.0	116	1	NGF_NAJAT
18	20	100.0	119	1	ACPS_BACND
19	20	100.0	121	1	RK14_ODOSI
20	20	100.0	125	1	DHSC_COXBU
21	20	100.0	129	1	DHSC_COXBU
22	20	100.0	131	1	FABA_ECOLI
23	20	100.0	131	1	FABA_BOVIN
24	20	100.0	145	1	YPH2_MYCCA
25	20	100.0	146	1	R27B_ARATH
26	20	100.0	146	1	R27C_ARATH
27	20	100.0	146	1	RL15_BACSU
28	20	100.0	147	1	RL15_STAM
29	20	100.0	147	1	RL2A_HUMAN
30	20	100.0	147	1	RL2A_MOUSE
31	20	100.0	147	1	RL2A_MOUSE
32	20	100.0	148	1	R28A_SCHPO
33	20	100.0	150	1	R28B_SCHPO
					P5728 schizosacch
					P33101 micrococcus

34	20	100.0	151	1	RL15_STRCO	P46787 streptomyc
35	20	100.0	151	1	TCPR_VIBCH	P29483 vibrio chol
36	20	100.0	151	1	YABR_HAEIN	P45056 haemophilus
37	20	100.0	152	1	TR13_HUMAN	Q15649 homo sapien
38	20	100.0	152	1	YABR_ECOLI	P22186 escherichia
39	20	100.0	153	1	NRDI_MYCPNA	P43961 mycoplasma
40	20	100.0	160	1	Y178_HAEIN	P43961 haemophilus
41	20	100.0	160	1	IL3_RAT	P04823 rattus norv
42	20	100.0	168	1	YWV1_CAEEL	Q11088 caenorhabdi
43	20	100.0	170	1	YFIR_BACND	O9867 bacillus ha
44	20	100.0	172	1	YFIR_BACND	P76597 escherichia
45	20	100.0	173	1	DEF_BUCAT	P57563 buchnera ap

## ALIGNMENTS

RESULT 1  
TERM\_BPM2  
ID TERM\_BPM2 STANDARD: PRT: 15 AA.  
AC P19897;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA terminal protein (Protein GP3) (Fragment).  
GN 3 OR F.  
OS Bacteriophage M2.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.  
OX NCBI\_TaxID=10751;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90128268; PubMed=2515115;  
RA Matsumoto K., Takano H., Kim C.I., Hirokawa H.;

\*Primary structure of bacteriophage M2 DNA polymerase: conserved segments within protein-priming DNA polymerases and DNA polymerase I or Escherichia coli.;

Gene 84:247-255(1989).  
-I- FUNCTION: DNA TERMINAL PROTEIN IS LINKED TO THE 5' ENDS OF BOTH STRANDS OF THE GENOME THROUGH A PHOSPHODIESTER BOND BETWEEN THE BETA-HYDROXYL GROUP OF A SERINE RESIDUE AND THE 5'-PHOSPHATE OF THE TERMINAL DEOXYRIBONATE. THIS PROTEIN IS ESSENTIAL FOR DNA REPLICATION AND IS INVOLVED IN THE PRIMING OF DNA ELONGATION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: M33144; AAA32367.1; -.  
DR PIR: PQ0017; PQ0017.  
KW Early protein; DNA replication; DNA priming;  
KW Covalent protein-DNA linkage.  
FT NON\_TER 1  
FT SITE 5 7  
SQ SEQUENCE 15 AA: 1797 MW: D3CBAF8759DEA06 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 VNLD 4  
1111  
Db 8 VNLD 11

RESULT 2  
CSP4\_STRMT STANDARD: PRT: 40 AA.  
ID CSP4\_STRMT  
AC O33675;  
DT 30-MAY-2000 (Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Competence stimulating peptide precursor (CSP).
GN COMC.
OS Streptococcus mitis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=28037;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 12261;
RX MEDLINE=98012953; PubMed=9352904;
RA Haavarstein L.S., Hakenbeck R., Gaustad P.;
RT "Natural competence in the genus Streptococcus: evidence that
RT streptococci can change phenotype by interspecies recombinational
RT exchanges."
RL J. Bacteriol. 179:6589-6594(1997).
CC -I- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
CC FOR GENETIC TRANSFORMATION.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE COMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ000875; CA04365.1; -
DR InterPro: IPR004288; COMC.
DR Pfam: PF03047; COMC. 1.
KW Pheromone; Competence.
FT PROPEP 1 24
FT CHAIN 25 40
FT SEQUENCE 40 AA; 4898 MW; 599DC7A3BADA1822 CRC64;
SQ
Query Match 100.0%; Score 20; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
DB 5 VNLD 8

RESULT 3
YDBF_SCHPO
ID YDBF_SCHPO STANDARD; PRT; 59 AA.
AC Q10365;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 6.6 kDa protein C22E12.15 in chromosome I.
GN SPAC22E12.15.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z70043; CA93902.1; -
DR Hypothetical protein
KW SEQUENCE 59 AA; 6594 MW; 39DA9EA0C4168549 CRC64;
SQ
Query Match 100.0%; Score 20; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
DB 19 VNLD 22

RESULT 4
ANPL_AUSBR
ID ANPL_AUSBR STANDARD; PRT; 63 AA.
AC P12100;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Antifreeze peptide ABL.
OS Austrolachnys brachycephalus (Antarctic eel pout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Zoarcoidei;
OC Zoarcidae; Austrolachnys.
OX NCBI_TaxID=8195;
RN [1]
RP SEQUENCE.
RX MEDLINE=89323219; PubMed=2752054;
RA Cheng C.-H.C., Devries A.L.;
RT "Structures of antifreeze peptides from the antarctic eel pout,
RT Austrolachnys brachycephalus."
RT Biochim. Biophys. Acta 997:55-64(1989).
CC -I- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT BY
CC ABSORBING ICE AND INHIBITING ITS GROWTH.
CC -I- SIMILARITY: BELONGS TO THE TYPE-III AFP FAMILY.
CC -----
DR PIR: S04973; S04973.
DR HSP: P35753; 3RDN.
DR InterPro: IPR000029; Antifreeze_type_III.
DR Pfam: PF01354; Antifreeze_1.
DR PRINTS: PR00357; ANTIFREEZE211.
DR Prodom: PD003258; Antifreeze_type_III; 1.
KW Antifreeze protein.
FT SITE 8 8 IMPORTANT FOR ICE-BINDING
FT SITE 13 13 (BY SIMILARITY).
FT SITE 17 17 IMPORTANT FOR ICE-BINDING
FT SITE 17 17 (BY SIMILARITY).
FT SITE 43 43 IMPORTANT FOR ICE-BINDING
FT SITE 43 43 (BY SIMILARITY).
SQ SEQUENCE 63 AA; 6646 MW; AC84FD1424719384 CRC64;
Query Match 100.0%; Score 20; DB 1; Length 63;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
DB 48 VNLD 51

RESULT 5
RGS3_RAT
ID RGS3_RAT STANDARD; PRT; 67 AA.
AC P49797;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Regulator of G-protein signaling 3 (RGS3) (Fragment).

```

```

GN RGS3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=96140645; PubMed=8548815;
RA Koelle M.R., Horvitz H.R.;
RT "Egr-10 regulates G protein signaling in the C. elegans nervous
RT system and shares a conserved domain with many mammalian proteins."
RL Cell 84:115-125(1996).
RN [2]
RP PHOSPHORYLATION.
RX MEDLINE=20167219; PubMed=10702309;
RA Pedram A., Razandi M., Kehrl J., Levin E.R.;
RT "Natriuretic peptides inhibit G protein activation. Mediation through
RT cross-talk between cyclic GMP-dependent protein kinase and regulators
RT of G protein-signaling proteins."
RL J. Biol. Chem. 275:7365-7372(2000).
CC -1- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC THEIR INACTIVE GDP-BOUND FORM.
CC -1- PHOSPHORYLATED BY CYCLIC GMP-DEPENDENT PROTEIN KINASE.
CC -1- SIMILARITY: CONTAINS 1 RGS DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U32434; AAC52371.1; -.
DR HSRP: P49799; IAGR.
DR InterPro: IPR000342; RGS.
DR Pfam: PF00615; RGS; 1.
DR ProDom: PD001580; RGS; 1.
DR PROSITE: PS00132; RGS; 1.
KW Signal transduction inhibitor; Phosphorylation.
FT NON_TER 1
FT DOMAIN 1
FT DON_TER <1 >67 RGS.
FT NON_TER 67
FT SQ SEQUENCE 67 AA; 7641 MW; 06FE7630E1CE3AF5 CRC64;
Query Match 100.0%; Score 20; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VILD 4
Db 35 VILD 38

```

```

RT "Cloning and molecular characterization of the secy genes from
RT Bacillus licheniformis and Staphylococcus carnosus: comparative
RT analysis of nine members of the Secy family.";
RL Mol. Gen. Genet. 235:147-152(1992).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X70087; -; NOT_ANNOTATED_CDS.
DR PIR: S34404; S34404.
DR InterPro: IPR001196; Ribosomal_L15.
DR Pfam: PF00256; L15; 1.
DR PROSITE: PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein; rRNA-binding.
FT NON_TER 1
FT SQ SEQUENCE 72 AA; 7441 MW; 6DC3BA681492E6C9 CRC64;
Query Match 100.0%; Score 20; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VILD 4
Db 3 VILD 6

```

RESULT 7

```

ECLH_BOMMO
ID ECLH_BOMMO STANDARD; PRT; 88 AA.
AC P25331;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Eclosion hormone precursor (Ecdisis activator) (EH).
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92134263; PubMed=1370883;
RA Kamito T., Tanaka H., Sato B., Nagasawa H., Suzuki A.;
RT "Nucleotide sequence of cDNA for the eclosion hormone of the
RT silkworm, Bombyx mori, and the expression in a brain.";
RL Biochem. Biophys. Res. Commun. 182:514-519(1992).
RN [2]
RP SEQUENCE OF 27-87.
RA Kono T., Nagasawa H., Isogai A., Fugo H., Suzuki A.;
RT "Amino acid sequence of eclosion hormone of the silkworm, Bombyx
RT mori.";
RL Agric. Biol. Chem. 51:2307-2308(1987).
CC -1- FUNCTION: NEUROPEPTIDE THAT TRIGGERS THE PERFORMANCE OF ECDYSIS
CC BEHAVIORS AT THE END OF A MOLT. IT TRIGGERS ADULT BEHAVIOR
CC PATTERNS: LARVAL, PUPAL AND ADULT ECDYSIS, AND PLASTICIZATION
CC DURING THE MOLT.
CC -1- SIMILARITY: HIGH, TO OTHER INSECTS ECLOSION HORMONE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```

DR EMBL: D10135; BAA01012.1; -.
DR PIR: JS0644; JS0644.
KW Hormone; Neuropeptide; Signal.
FT SIGNAL 1 26
FT CHAIN 27 88 ECLOSION HORMONE.
FT DISULFID 40 64 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 47 75 BY SIMILARITY.
FT DISULFID 31 31 A -> S (IN REF. 2).
FT CONFLICT 33 33 S -> G (IN REF. 2).
FT CONFLICT 36 46 Q -> E (IN REF. 2).
FT CONFLICT 67 68 AR -> FK (IN REF. 2).
FT CONFLICT 70 71 KD -> DL (IN REF. 2).
FT CONFLICT 71 71
SQ SEQUENCE 88 AA; 9505 MW; 429DDCOADDAAD852 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNUD 4
DB 19 VNUD 22

RESULT 8
EF1B_SUISO STANDARD; PRT; 90 AA.
ID EF1B_SUISO
AC Q64214;
DR 15-JUL-1998 (Rel. 36, Created)
DR 30-MAY-2000 (Rel. 39, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor 1-beta (EF-1-beta) (aEF-1beta).
GN EF1B OR SS05345.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11 AND 82-90.
RC STRAIN=DSM 5833 / Mt-4;
RX MEDLINE=95359209; PubMed=7632739;
RA Arcari P., Raimo G., Iannicello G., Gallo M., Bocchini V.;
RT "The first nucleotide sequence of an archaeal elongation factor 1
beta gene.";
RL Biochim. Biophys. Acta 1263:86-88(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault G., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RN [3]
RP SEQUENCE, FUNCTION, SUBUNITS, AND MASS SPECTROMETRY.
RC STRAIN=DSM 5833 / Mt-4;
RX MEDLINE=96186282; PubMed=8652615;
RA Raimo G., Masullo M., Savino G., Searano G., Iannicello G.,
RA Parente A., Bocchini V.;
RT "Archaeal elongation factor 1 beta is a dimer. Primary structure,
molecular and biochemical properties.";
RL Biochim. Biophys. Acta 1293:106-112(1996).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF GDP FOR GTP IN EF-1-ALPHA/GDP,
THUS ALLOWING THE REGENERATION OF EF-1-ALPHA/GTP THAT COULD THEN
BE USED TO FORM THE TERNARY COMPLEX EF-1-ALPHA/GTP/AA/TRNA.
CC -1- SUBUNIT: HOMODIMER.
CC -1- MASS SPECTROMETRY: MW=10006; METHOD=Electrospray.
CC -1- SIMILARITY: BELONGS TO THE EF-1-BETA/EF-1-DELTA FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL: X76768; CAA54164.1; -.
DR InterPro: IPR001326; EF1BD.
DR Pfam: PF00736; EF1BD; 1.
KW Elongation factor; Protein biosynthesis; Complete proteome.
FT INIT_MBT 0
FT SEQUENCE 90 AA; 10005 MW; 24CB357FD3259C8A CRC64;

Query Match 100.0%; Score 20; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNUD 4
DB 16 VNUD 19

RESULT 9
EF1B_PYRAB STANDARD; PRT; 91 AA.
ID EF1B_PYRAB
AC Q9VZP6;
DR 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor 1-beta (EF-1-beta) (aEF-1beta).
GN EF1B OR PAB3009.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
OX NCBI_TaxID=22992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Helling R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF GDP FOR GTP IN EF-1-ALPHA/GDP,
THUS ALLOWING THE REGENERATION OF EF-1-ALPHA/GTP THAT COULD THEN
BE USED TO FORM THE TERNARY COMPLEX EF-1-ALPHA/GTP/AA/TRNA (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EF-1-BETA/EF-1-DELTA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL: AJ248283; CAB48952.1; -.
DR InterPro: IPR001326; EF1BD.
DR Pfam: PF00736; EF1BD; 1.
KW Elongation factor; Protein biosynthesis; Complete proteome.
SQ SEQUENCE 91 AA; 10239 MW; 45A032D81D54F1EA CRC64;

Query Match 100.0%; Score 20; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNUD 4
DB 19 VNUD 22

```



```

RESULT 10
RL15_STRSC          STANDARD:          PRT:          96 AA.
ID   RL15_STRSC
AC   P43415;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DE   50S ribosomal protein L15 (fragment).
GN   rplO.
OS   Streptomyces scabies.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1930;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FL1;
RX   MEDLINE=96001249; PubMed=7557484;
RA   Hale V.A., O'Brien I., Schottel J.L.;
RT   "Cloning and sequencing of a secY homolog from Streptomyces scabies.";
RL   Gene 163:87-92(1995).
CC   -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA (BY
CC   SIMILARITY).
CC   -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U19606; AA85556.1; -
DR   InterPro; IPR001196; Ribosomal_L15.
DR   Pfam; PF00256; L15; 1.
DR   Prosite; PS00475; RIBOSOMAL_L15; 1.
KW   Ribosomal protein; rRNA-binding.
FT   NON_TER
SQ   SEQUENCE 96 AA; 10075 MW; 0BAAC499331BF70E CRC64;

Query Match          100.0%; Score 20; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIND 4
    |||
Db 26 VIND 29

RESULT 11
SR19_ARCFU          STANDARD:          PRT:          104 AA.
ID   SR19_ARCFU
AC   O29010;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DE   Signal recognition particle 19 kDa protein (SRP19).
GN   SRP19 OR AF1258.
OS   Archaeoglobus fulgidus.
OC   Archaeae; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Fleischmann R.D., Kerlavage A.R., Graham D.E., Kyprides N.C.,
RA   Kikness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodok A., Zhou L.,

```

```

RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=20150251; PubMed=10684931;
RA   Bhuiyan S.H., Gowda K., Hotokezaka H., Zwieb C.;
RT   "Assembly of archaeal signal recognition particle from recombinant
RT   components.";
RL   Nucleic Acids Res. 28:1365-1373(2000).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=20496765; PubMed=11041851;
RA   Diener J.L., Wilson C.;
RT   "Role of SRP19 in assembly of the Archaeoglobus fulgidus signal
RT   recognition particle.";
RL   Biochemistry 39:12862-12874(2000).
CC   -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY, BINDS DIRECTLY
CC   TO 7S RNA AND MEDIATES BINDING OF THE 54 KDA SUBUNIT OF THE SRP.
CC   -1- SUBUNIT: ARCHAEOAL SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA
CC   MOLECULE OF 300 NUCLEOTIDES AND TWO PROTEIN SUBUNITS. SRP54 AND
CC   SRP19.
CC   -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -1- SIMILARITY: BELONGS TO THE SRP19 FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AE001017; AAB89988.1; -
DR   TIGR; AF1258; -
DR   InterPro; IPR002778; SRP19.
DR   Pfam; PF01922; SRP19; 1.
DR   Prodom; PD006609; SRP19; 1.
KW   Signal recognition particle; RNA-binding; Ribonucleoprotein;
KW   Complete proteome.
SQ   SEQUENCE 104 AA; 12405 MW; 72D5DDABB4E89E64 CRC64;

Query Match          100.0%; Score 20; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIND 4
    |||
Db 9 VIND 12

RESULT 12
HAL2_CANAL          STANDARD:          PRT:          105 AA.
ID   HAL2_CANAL
AC   P46594;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DE   Halotolerance protein HAL2 (Fragment).
GN   HAL2 OR MET22.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WO-1;
RX   MEDLINE=95242840; PubMed=7725800;

```

RA Sychorova H., Souciet J.-L.;  
 RT "CAN1, a gene encoding a permease for basic amino acids in Candida  
 RT albicans.";  
 RL Yeast 10:1647-1651(1994).  
 CC -1- FUNCTION: INVOLVED IN SALT TOLERANCE AND IN METHIONINE  
 CC BIOSYNTHESIS (BY SIMILARITY).  
 CC -1- PATHWAY: METHIONINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X76689; -; NOT\_ANNOTATED\_CDS.  
 DR HSSP: P29218; 1MB.  
 DR InterPro: IPR000760; Inositol\_P.  
 DR Pfam: PF00459; Inositol\_P: 1.  
 DR PROSITE: PS00629; IMP\_1; PARTIAL.  
 DR PROSITE: PS00630; IMP\_2; 1.  
 KW Methionine biosynthesis.  
 FT NON\_TER  
 SO SEQUENCE 105 AA; 11338 MW; EB6938DD4F262F1 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 92;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 UNLD 4  
 Db 9 UNLD 12  
 Db 9 UNLD 12  
 RESULT 13  
 ELIE\_BACSU STANDARD; PRT; 106 AA.  
 ID ELIE\_BACSU  
 AC P24502;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Flagellar hook-basal body complex protein flie.  
 GN FLIE.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91285431; PubMed-1905667;  
 RA Zuberi A.R., Ying C., Bischoff D.S., Ordal G.W.;  
 RT "Gene-protein relationships in the flagellar hook-basal body complex  
 RT of Bacillus subtilis: sequences of the fligB, fligC, fligG, flie and  
 RT flif genes.";  
 RL Gene 101:23-31(1991).  
 CC -1- SIMILARITY: BELONGS TO THE FLIE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M54965; AAA22443.1; -;  
 DR EMBL: Z99112; CAB13493.1; -;  
 DR PIR: JG0021; JG0021.  
 DR Subtilist; BG10239; flie.  
 DR InterPro: IPR001624; flie.  
 DR Pfam: PF02049; flie; 1.

DR PRINTS: PR01006; FLGHOOKFLIE.  
 KW Flagella: Complete proteome.  
 SO SEQUENCE 106 AA; 11577 MW; 7A3F1B7FE653AC9C CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 UNLD 4  
 Db 68 UNLD 71  
 Db 68 UNLD 71  
 RESULT 14  
 SPEH\_ARCFU STANDARD; PRT; 112 AA.  
 ID SPEH\_ARCFU  
 AC 028663;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdometDC)  
 DE (SampC) [Constraints: S-adenosylmethionine decarboxylase beta chain; S-  
 DE adenosylmethionine decarboxylase alpha chain].  
 GN SPEH OR AF1610.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 MEDLINE-98049343; PubMed=9389475;  
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougeny B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöck A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -1- FUNCTION: DECARBOXYLATION OF S-ADENOSYLMETHIONINE PROVIDES THE  
 CC AMINOACIDIC MOIETY REQUIRED FOR SPERMIDINE BIOSYNTHESIS FROM  
 CC PUTRESCINE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: S'-adenosyl-L-methionine = (5'-deoxy-5'-  
 CC adenosyl)(3'-aminopropyl) methylsulfonium salt + CO(2).  
 CC -1- COFACTOR: Pyruvoyl group (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC ADOMETDC FAMILY. SUBFAMILY  
 CC 1  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000991; AAB89640.1; -;  
 DR TIGR: AF1610; -;  
 DR InterPro: IPR003826; DUF206.  
 KW Spermidine biosynthesis; Lyase; Decarboxylase; Lysogen; Pyruvate;  
 KW Complete proteome.  
 FT CHAIN 1 61 S-ADENOSYLMETHIONINE DECARBOXYLASE BETA  
 FT CHAIN 62 112 CHAIN (BY SIMILARITY).  
 FT SITE 61 62 S-ADENOSYLMETHIONINE DECARBOXYLASE ALPHA  
 FT SITE 61 62 CHAIN (BY SIMILARITY).  
 FT SITE 61 62 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).

FT MOD\_RES 62 62 CONVERTED TO A PYRUVYL GROUP (BY  
FT SEQUENCE 112 AA; 12733 MW; 95B250A1306D2F30 CRC64;  
SQ  
Query Match 100.0%; Score 20; DB 1; Length 112;  
Best Local Similarity 100.0%; Pred. NO. 99;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 40 VNLD 43  
Search completed: October 30, 2002, 12:27:57  
Job time : 1.34398 secs

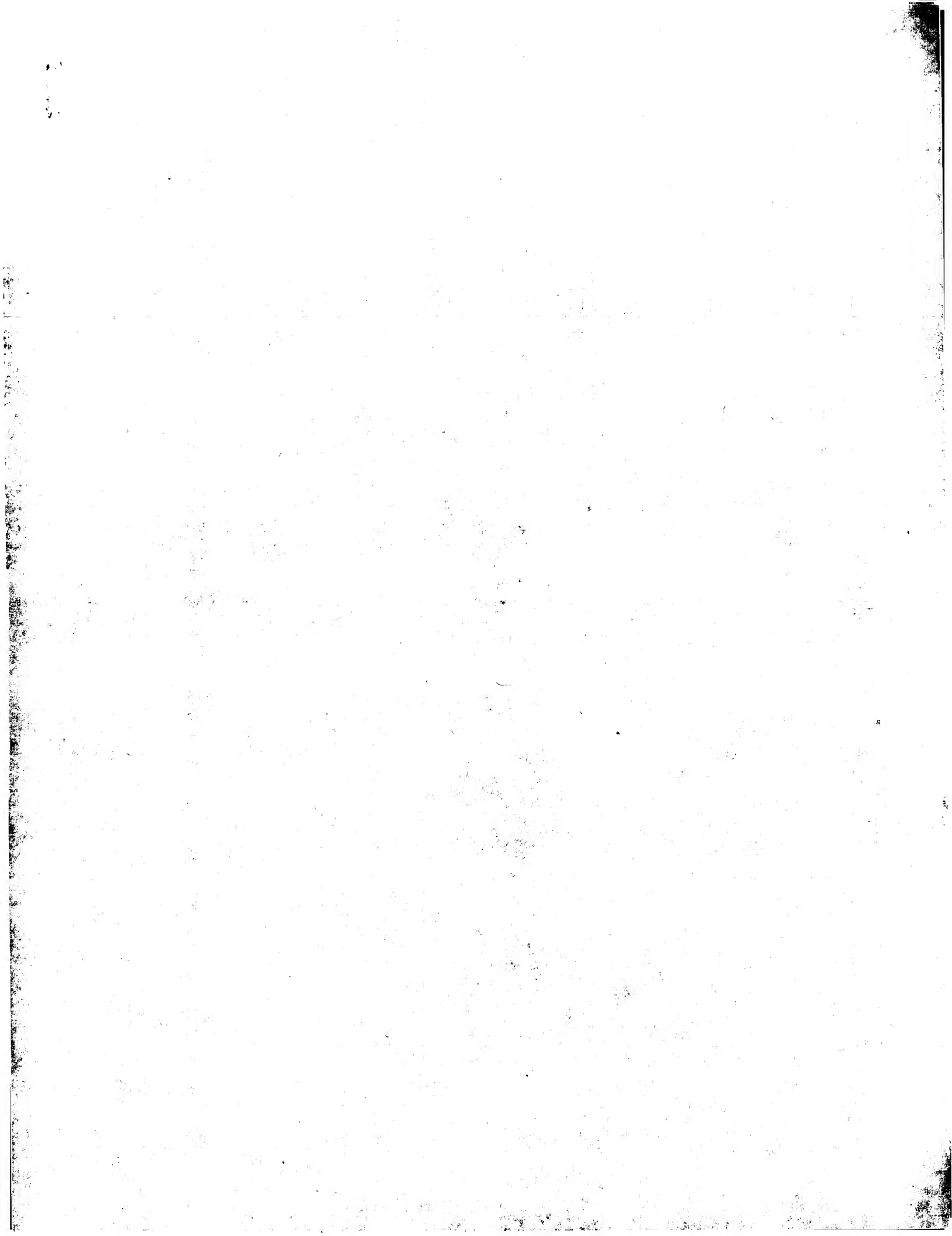
OY 1 VNLD 4  
Db 75 VNLD 78

RESULT 15

NGF\_MAJAT  
ID NGF\_MAJAT STANDARD; PRT; 116 AA.  
AC P21377;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE Nerve growth factor (NGF).  
OS Naja atra (Chinese cobra), and  
OS Naja naja kaouthia (Monocled cobra) (Naja naja siamensis).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Lepidodactylus; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Elapidae; Elapinae; Naja.  
OX NCBI\_TaxID=8656, 8649;  
RN (1)  
RP SEQUENCE.  
RC SPECIES=N.n.kaouthia; TISSUE=Venom;  
RX MEDLINE=90147847; PubMed=2619756;  
RA Oda T., Ohta M., Inoue S., Ikeda K., Furukawa S., Hayashi K.;  
RT "Amino acid sequence of nerve growth factor purified from the venom  
of the Formosan cobra Naja naja atra.";  
RL Biochem. Int. 19:909-917(1989).  
RN (2)  
RP SEQUENCE.  
RC SPECIES=N.n.kaouthia; TISSUE=Venom;  
RX MEDLINE=91138755; PubMed=1995338;  
RA Inoue S., Oda T., Koyama J., Ikeda K., Hayashi K.;  
RT "Amino acid sequences of nerve growth factors derived from cobra  
venoms.";  
RL FEBS Lett. 279:38-40(1991).  
CC -!- FUNCTION: NERVE GROWTH FACTOR IS IMPORTANT FOR THE DEVELOPMENT AND  
CC MAINTENANCE OF THE SYMPATHETIC AND SENSORY NERVOUS SYSTEMS. IT  
CC STIMULATES DIVISION AND DIFFERENTIATION OF SYMPATHETIC AND  
CC EMERGENCY SENSORY NEURONS AS WELL AS BASAL FOREBRAIN CHOLINERGIC  
CC NEURONS IN THE BRAIN.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.  
CC PIR: S13965, S13965.  
DR HSPF, P01139, ILET.  
DR InterPro: IPR002400; GE\_cysknot.  
DR InterPro: IPR002072; NGF.  
DR Pfam: PF00243; NGF; 1.  
DR PRINTS: PRO0438; GFCYSKNOT.  
DR PRINTS: PRO0268; NGF.  
DR PRODOM: PRO02052; NGF; 1.  
DR SMART: SM00140; NGF; 1.  
DR PROSITE: PS00248; NGF\_1; 1.  
DR PROSITE: PS50270; NGF\_2; 1.  
KW Growth factor.  
FT DISULFID 14 78 BY SIMILARITY.  
FT DISULFID 56 106 BY SIMILARITY.  
FT DISULFID 66 108 BY SIMILARITY.  
SQ SEQUENCE 116 AA; 13064 MW; DAB35421093F3B06 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 116;  
Best Local Similarity 100.0%; Pred. NO. 1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
Db 75 VNLD 78



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 1.24816 Seconds

(without alignments)  
554,401 Million cell updates/sec

Title: US-09-724-571-104

Perfect score: 20

Sequence: 1 VNLD 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	32	2 09R507	09R507 aeromonas h
2	20	100.0	33	12 089871	089871 budgerigar
3	20	100.0	33	12 089872	089872 budgerigar
4	20	100.0	33	12 089873	089873 budgerigar
5	20	100.0	33	12 089874	089874 budgerigar
6	20	100.0	33	12 089875	089875 budgerigar
7	20	100.0	33	12 089876	089876 budgerigar
8	20	100.0	33	12 089877	089877 budgerigar
9	20	100.0	33	12 089878	089878 budgerigar
10	20	100.0	33	12 089879	089879 budgerigar
11	20	100.0	33	12 089880	089880 budgerigar
12	20	100.0	33	12 089881	089881 budgerigar
13	20	100.0	33	12 089882	089882 budgerigar
14	20	100.0	33	12 089883	089883 budgerigar
15	20	100.0	33	12 089884	089884 budgerigar
16	20	100.0	33	12 089885	089885 budgerigar

17	20	100.0	33	12 089886	089886 budgerigar
18	20	100.0	33	12 089887	089887 budgerigar
19	20	100.0	33	16 09PH79	09PH79 xylella fas
20	20	100.0	33	16 09RJ3	09RJ3 streptococ
21	20	100.0	40	2 068327	068327 caudobacter
22	20	100.0	41	2 09R2X2	09R2X2 borrellia bu
23	20	100.0	44	2 09SOL7	09SOL7 streptomyc
24	20	100.0	45	4 09HCW4	09HCW4 homo sapien
25	20	100.0	49	5 09VGM9	09VGM9 drosophila
26	20	100.0	49	7 09G1Q2	09G1Q2 eschrichtiu
27	20	100.0	49	8 020180	020180 chlorella v
28	20	100.0	49	13 09PVR4	09PVR4 gallus gall
29	20	100.0	51	2 051106	051106 neisseria m
30	20	100.0	61	2 056859	056859 yersinia en
31	20	100.0	62	10 039072	039072 arbidopsi
32	20	100.0	67	6 09TU51	09TU51 sus scrofa
33	20	100.0	67	13 09IB58	09IB58 xenopus lae
34	20	100.0	67	13 09IB55	09IB55 xenopus lae
35	20	100.0	76	2 09AH64	09AH64 neisseria g
36	20	100.0	79	2 096055	096055 salmonella
37	20	100.0	79	12 084524	084524 paramecium
38	20	100.0	80	16 099SP0	099SP0 staphylococ
39	20	100.0	82	7 09XR70	09XR70 rattus fusc
40	20	100.0	82	7 09XRT7	09XRT7 rattus fusc
41	20	100.0	82	15 090069	090069 human immun
42	20	100.0	83	2 09ZHA3	09ZHA3 morganella
43	20	100.0	83	2 09XA75	09XA75 streptomyc
44	20	100.0	84	13 09W717	09W717 gallus gall
45	20	100.0	84	16 053502	053502 mycobacteri

## ALIGNMENTS

## RESULT 1

ID 09R507 PRELIMINARY; PRT; 32 AA.  
AC 09R507;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE S-LAYER PROTEIN.  
OS Aeromonas hydrophila.  
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;  
OC Aeromonas.  
OX NCBI\_TaxID=644;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=92104988; PubMed=1370287;  
RA Kostrezyńska M., Docley J.S., Shimojo T., Sakata T., Trust T.J.,  
RT "Antigenic diversity of the S-layer proteins from pathogenic strains  
RT of Aeromonas hydrophila and Aeromonas veronii biotype sobria.";  
RL J. Bacteriol. 174:40-47(1992).  
SQ SEQUENCE 32 AA; 2931 MW; C1E8852A6DA2C8A5 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 32;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLD 4  
Db 1 VNLD 4

## RESULT 2

ID 089871 PRELIMINARY; PRT; 33 AA.  
AC 089871;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
DE LARGE T ANTIGEN (FRAGMENT).  
OS Budgerigar fledgling disease virus (BFDV).

```

OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN SEQUENCE FROM N.A.
RC STRAIN=MCFL97;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF054369; AAC33609.1; -.
FT NON_TER 1
FT NON_TER 33
SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 3
089872 PRELIMINARY; PRT; 33 AA.
AC 089872;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgetiglar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LBMI92;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF054370; AAC33610.1; -.
FT NON_TER 1
FT NON_TER 33
SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 4
089873 PRELIMINARY; PRT; 33 AA.
AC 089873;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgetiglar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GC4292;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF054371; AAC33611.1; -.
FT NON_TER 1
FT NON_TER 33
SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

```

```

SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 5
089874 PRELIMINARY; PRT; 33 AA.
AC 089874;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgetiglar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECLX91;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF054372; AAC33612.1; -.
FT NON_TER 1
FT NON_TER 33
SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 6
089875 PRELIMINARY; PRT; 33 AA.
AC 089875;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgetiglar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB85;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF054373; AAC33613.1; -.
FT NON_TER 1
FT NON_TER 33
SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

```

```
RESULT 7
089876 ID 089876 PRELIMINARY; PRT: 33 AA.
AC 089876;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgerigar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XCFL87;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
FT EMBL: AF054374; AAC33614.1; -
FT NON_TER 1 1
FT 33 33
SQ SEQUENCE 33 AA: 3713 MW: ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 8
089877 ID 089877 PRELIMINARY; PRT: 33 AA.
AC 089877;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgerigar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCIL88;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
FT EMBL: AF054375; AAC33615.1; -
FT NON_TER 1 1
FT 33 33
SQ SEQUENCE 33 AA: 3697 MW: 2034126F3BBFA88 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 9
089878 ID 089878 PRELIMINARY; PRT: 33 AA.
AC 089878;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgerigar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BDTX88;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
FT EMBL: AF054376; AAC33616.1; -
FT NON_TER 1 1
FT 33 33
SQ SEQUENCE 33 AA: 3697 MW: 2034126F3BBFA88 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 10
089879 ID 089879 PRELIMINARY; PRT: 33 AA.
AC 089879;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgerigar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RNA91, LRTX93;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
FT EMBL: AF054377; AAC33617.1; -
FT NON_TER 1 1
FT 33 33
SQ SEQUENCE 33 AA: 3713 MW: ABA4126F3BBFAA79 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4
Db 27 VNLD 30

RESULT 11
089880 ID 089880 PRELIMINARY; PRT: 33 AA.
AC 089880;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases)
DE LARGE T ANTIGEN (FRAGMENT).
OS Budgerigar fledgling disease virus (BFDV).
OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.
OX NCBI_TaxID=10625;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BDTX89;
RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;
RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
FT EMBL: AF054378; AAC33618.1; -
FT NON_TER 1 1
FT 33 33
SQ SEQUENCE 33 AA: 3713 MW: ABA4126F3BBFAA79 CRC64;
```

Query Match 100.0%; Score 20; DB 12; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 27 VNLD 30

RESULT 12  
 ID 089881 PRELIMINARY; PRT: 33 AA.  
 AC 089881;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE LARGE T ANTIGEN (FRAGMENT).  
 OS Budgerigar fledgling disease virus (BFDV).  
 OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.  
 OX NCBI\_TaxID=10625;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BDGA81-A;  
 RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;  
 RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus."  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF054379; AAC33619.1; -  
 FT NON\_TER 1 1  
 FT 33 33  
 SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 27 VNLD 30

RESULT 13  
 ID 089882 PRELIMINARY; PRT: 33 AA.  
 AC 089882;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE LARGE T ANTIGEN (FRAGMENT).  
 OS Budgerigar fledgling disease virus (BFDV).  
 OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.  
 OX NCBI\_TaxID=10625;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BDGA81-B;  
 RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;  
 RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus."  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF054380; AAC33620.1; -  
 FT NON\_TER 1 1  
 FT 33 33  
 SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 27 VNLD 30

RESULT 14  
 C93883

ID 089883 PRELIMINARY; PRT: 33 AA.  
 AC 089883;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE LARGE T ANTIGEN (FRAGMENT).  
 OS Budgerigar fledgling disease virus (BFDV).  
 OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.  
 OX NCBI\_TaxID=10625;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BCEFL92;  
 RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;  
 RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus."  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF054381; AAC33621.1; -  
 FT NON\_TER 1 1  
 FT 33 33  
 SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 27 VNLD 30

RESULT 15  
 ID 089884 PRELIMINARY; PRT: 33 AA.  
 AC 089884;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE LARGE T ANTIGEN (FRAGMENT).  
 OS Budgerigar fledgling disease virus (BFDV).  
 OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.  
 OX NCBI\_TaxID=10625;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ECEFL91;  
 RA Phalen D.N., Wilson V.G., Gaskin J.M., Derr J.N., Graham D.L.;  
 RT "Genetic Diversity in 20 Variants of the Avian Polyomavirus."  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF054382; AAC33622.1; -  
 FT NON\_TER 1 1  
 FT 33 33  
 SQ SEQUENCE 33 AA; 3713 MW; ABA4126F3BBFAA79 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLD 4  
 ||||  
 DB 27 VNLD 30

Search completed: October 30, 2002, 12:30:17  
 Job time: 3.24816 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 ; Search time 51.4447 Seconds

(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-75

Perfect score: 1907  
Sequence: 1 MVDNLKRGSGQGYVEMTVG.....VVFDRARKRIGFAVSACHVH 361

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1901	99.7	501	US-09-009-191-2	Sequence 2, Appl1
2	1890	99.1	774	US-09-009-191-4	Sequence 4, Appl1
3	1106	58.0	518	US-08-999-723-2	Sequence 2, Appl1
4	1106	58.0	518	US-09-434-427-2	Sequence 2, Appl1
5	1088	57.1	514	US-09-717-432-2	Sequence 2, Appl1
6	1088	57.1	514	US-09-912-484-2	Sequence 2, Appl1
7	298.5	15.7	396	US-08-208-007A-13	Sequence 13, Appl1
8	298.5	15.7	396	US-09-032-523-9	Sequence 9, Appl1
9	295.5	15.5	412	US-08-208-007A-12	Sequence 12, Appl1
10	295.5	15.5	412	US-08-974-691-4	Sequence 4, Appl1
11	279.5	14.7	458	5217891-15	Patent No. 5217891
12	273.5	14.3	409	US-09-640-305-6	Sequence 6, Appl1
13	273.5	14.3	409	US-08-360-673-6	Sequence 6, Appl1
14	273.5	14.3	427	US-08-846-021A-8	Sequence 8, Appl1
15	270	14.2	410	US-08-088-633-2	Sequence 2, Appl1
16	270	14.2	410	US-08-245-756-2	Sequence 2, Appl1
17	270	14.2	410	US-08-441-750-2	Sequence 2, Appl1
18	270	14.2	410	US-08-441-751-2	Sequence 2, Appl1
19	270	14.2	410	PCT-US92-02521-2	Sequence 2, Appl1
20	249	13.1	349	US-09-032-523-3	Sequence 3, Appl1
21	227	11.9	398	US-08-328-314-2	Sequence 2, Appl1
22	227	11.9	398	US-08-731-045-2	Sequence 2, Appl1
23	212	11.1	419	US-08-974-691-3	Sequence 3, Appl1
24	210	11.0	397	US-09-079-415-2	Sequence 2, Appl1
25	195.5	10.3	427	US-07-958-222A-2	Sequence 2, Appl1
26	194.5	10.2	430	US-08-535-237-2	Sequence 2, Appl1
27	193	10.1	330	US-08-115-753-1	Sequence 1, Appl1

#### ALIGNMENTS

28	193	10.1	419	3	US-08-115-753-2	Sequence 2, Appl1
29	193	10.1	419	3	US-08-115-753-33	Sequence 33, Appl1
30	184	9.6	445	4	US-08-974-691-6	Sequence 6, Appl1
31	184	9.6	451	4	US-08-974-691-2	Sequence 2, Appl1
32	180	9.4	420	4	US-09-008-227A-4	Sequence 4, Appl1
33	180	9.4	420	4	US-08-974-691-8	Sequence 8, Appl1
34	177.5	9.3	395	1	US-08-723-938-3	Sequence 3, Appl1
35	177.5	9.3	395	1	US-09-080-538-3	Sequence 3, Appl1
36	149	7.8	437	4	US-09-211-631-13	Sequence 13, Appl1
37	129.5	6.8	140	3	US-09-353-332-2	Sequence 2, Appl1
38	129.5	6.8	140	4	US-09-265-628-13	Sequence 13, Appl1
39	129.5	6.8	140	4	US-09-532-803-6	Sequence 6, Appl1
40	129.5	6.8	140	4	US-09-001-141-11	Sequence 11, Appl1
41	129.5	6.8	140	4	US-09-653-403-14	Sequence 14, Appl1
42	97	5.1	1030	4	US-09-091-117-2	Sequence 2, Appl1
43	95.5	5.0	280	4	US-09-160-246-14	Sequence 14, Appl1
44	85	4.5	377	2	US-08-853-659A-41	Sequence 41, Appl1
45	84.5	4.4	1097	2	US-08-680-326-39	Sequence 39, Appl1

#### RESULT 1

US-09-009-191-2  
Sequence 2, Application US/09009191

Patent No. 6319689

GENERAL INFORMATION:

APPLICANT: POWELL, DAVID

APPLICANT: CHAPMAN, CONRAD

APPLICANT: MURPHY, KAY

APPLICANT: SMITH, TRUDI

TITLE OF INVENTION: ASP2

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESS: RATNER & PRESTIA

STREET: P.O. BOX 980

CITY: VALLEY FORGE

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/009,191

FILING DATE: 20-JAN-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: UK 9701684.4

FILING DATE: 128-JAN-1997

ATTORNEY/AGENT INFORMATION:

NAME: PRESTIA, PAUL F

REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: GH-70368

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701

TELEX: 846169

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 501 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-009-191-2

Query Match 99.7%; Score 1901; DB 4; Length 501;

Best Local Similarity 99.7%; Pred. No. 1.4e-187;

Matches 360; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY      1 MVDNLGRKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAPHPFLHRYRQRLSSTYR 60
Db      63 MVDNLGRKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAPHPFLHRYRQRLSSTYR 122
QY      61 DLKRGVYVPYTGKWEGLGELTDLVSIPIHGPNTVVRANIAATFESDKFFINGSNWEGILGL 120
Db      123 DLKRGVYVPYTGKWEGLGELTDLVSIPIHGPNTVVRANIAATFESDKFFINGSNWEGILGL 182
QY      121 AYAEIARPDSDLPEPFDLSLVKQTHVPMILFSLQCGAGFPLNQEVLASVGSMTIGIDH 180
Db      183 AYAEIARPDSDLPEPFDLSLVKQTHVPMILFSLQCGAGFPLNQEVLASVGSMTIGIDH 242
QY      181 SLVYGLMYTPPIRREMYEVIIVRVEINGODLKMDCKEYNDKSIYDSGTTNLRPKKVF 240
Db      243 SLVYGLMYTPPIRREMYEVIIVRVEINGODLKMDCKEYNDKSIYDSGTTNLRPKKVF 302
QY      241 EAAVKSIAASSTKFPDGFMLGEOLVCMQAGTTPMNIFFVISLYLMGEVYNOSFRITIL 300
Db      303 EAAVKSIAASSTKFPDGFMLGEOLVCMQAGTTPMNIFFVISLYLMGEVYNOSFRITIL 362
QY      301 PQOYLRPVEDVATSDODCKKFAISQSSGTGYMGAVIMEGFYVFDRAKRRIIGFAVSACHV 360
Db      363 PQOYLRPVEDVATSDODCKKFAISQSSGTGYMGAVIMEGFYVFDRAKRRIIGFAVSACHV 422
QY      361 H 361
Db      423 H 423

```

```

RESULT 2
US-09-009-191-4
; Sequence 4, Application US/09009191
; Patent No. 6319689

```

```

; GENERAL INFORMATION:
; APPLICANT: POWELL, DAVID
; APPLICANT: CHAPMAN, CONRAD
; APPLICANT: MURPHY, KAY
; APPLICANT: SMITH, TRUDI
; TITLE OF INVENTION: ASP2
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,191
; FILING DATE: 20-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9701684.4
; FILING DATE: 28-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70368
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-009-191-4

```

```

Query Match          99.1%; Score 1890; DB 4; Length 774;
Best Local Similarity 99.2%; Pred. No. 3,7e-186;
Matches 358; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 MVDNLGRKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAPHPFLHRYRQRLSSTYR 60
Db      6 MVDNLGRKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAPHPFLHRYRQRLSSTYR 65
QY      61 DLKRGVYVPYTGKWEGLGELTDLVSIPIHGPNTVVRANIAATFESDKFFINGSNWEGILGL 120
Db      66 DLKRGVYVPYTGKWEGLGELTDLVSIPIHGPNTVVRANIAATFESDKFFINGSNWEGILGL 125
QY      121 AYAEIARPDSDLPEPFDLSLVKQTHVPMILFSLQCGAGFPLNQEVLASVGSMTIGIDH 180
Db      126 AYAEIARPDSDLPEPFDLSLVKQTHVPMILFSLQCGAGFPLNQEVLASVGSMTIGIDH 185
QY      181 SLVYGLMYTPPIRREMYEVIIVRVEINGODLKMDCKEYNDKSIYDSGTTNLRPKKVF 240
Db      186 SLVYGLMYTPPIRREMYEVIIVRVEINGODLKMDCKEYNDKSIYDSGTTNLRPKKVF 245
QY      241 EAAVKSIAASSTKFPDGFMLGEOLVCMQAGTTPMNIFFVISLYLMGEVYNOSFRITIL 300
Db      246 EAAVKSIAASSTKFPDGFMLGEOLVCMQAGTTPMNIFFVISLYLMGEVYNOSFRITIL 305
QY      301 PQOYLRPVEDVATSDODCKKFAISQSSGTGYMGAVIMEGFYVFDRAKRRIIGFAVSACHV 360
Db      306 PQOYLRPVEDVATSDODCKKFAISQSSGTGYMGAVIMEGFYVFDRAKRRIIGFAVSACHV 365
QY      361 H 361
Db      366 H 366

```

```

RESULT 3
US-08-999-723-2
; Sequence 2, Application US/08999723A
; Patent No. 6025180

```

```

; GENERAL INFORMATION:
; APPLICANT: Powell, David J.
; APPLICANT: Southan, Christopher
; APPLICANT: Chapman, Conrad G.
; APPLICANT: Pyans, Joanne R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH/0262
; CURRENT APPLICATION NUMBER: US/08/999,723A
; CURRENT FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-999-723-2

```

```

Query Match          58.0%; Score 1106; DB 3; Length 518;
Best Local Similarity 56.4%; Pred. No. 1.5e-105;
Matches 202; Conservative 60; Mismatches 92; Indels 4; Gaps 2;

```

```

QY      1 MVDNLGRKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAPHPFLHRYRQRLSSTYR 60
Db      80 MVDNLGRKSGGGYVEMTVGSPPTLNILVDTGSSNFVAGAAPHPFLHRYRQRLSSTYR 139
QY      61 DLKRGVYVPYTGKWEGLGELTDLVSIPIHGPNTVVRANIAATFESDKFFINGSNWEGILGL 120
Db      140 SKGFDVYVTKYTGSGWTFPGVEDLVITKGNFTSLVIVIAITFESSENFPLGKWNGLIGL 199
QY      121 AYAEIARPDSDLPEPFDLSLVKQTHVPMILFSLQCGAGFPLNQEVLASVGSMTIGIDH 180
Db      200 AYATLAKSSSLTFPDSLVTAQANIPNVFSWOMCGAGLPVAGS---GTNGSLVLGGLIEP 256

```

[illegible]

```

      1  APPLICANT: ZHU, YUAN
      2  APPLICANT: LI, XIAOTONG
      3  APPLICANT: CHRISTIE, GARY
      4  APPLICANT: POWELL, DAVID J.
      5  TITLE OF INVENTION: Mouse Aspartic Secretase-1 (masp1)
      6  FILE REFERENCE: GP-70663
      7  CURRENT APPLICATION NUMBER: US/09/717,432
      8  CURRENT FILING DATE: 2000-11-21
      9  PRIOR APPLICATION NUMBER: 60/166,974
     10  PRIOR FILING DATE: 1999-11-23
     11  NUMBER OF SEQ ID NOS: 2
     12  SOFTWARE: FastSeq for Windows Version 3.0
     13  SEQ ID NO: 2
     14  LENGTH: 514
     15  TYPE: PRT
     16  ORGANISM: MUS MUSCULUS
     17  US-09-717-432-2

Query Match      57.1%; Score 1088; DB 4; Length 514;
Best Local Similarity 55.9%; Pred. No. 1,1e-103;
Matches 200; Conservative 60; Mismatches 94; Indels 4; Gaps 2

QY      1  MYDNLKSGSGGCIYEMATYVSGSPQTLNLIVDTGSSNFVAGAAHPHPLHRYRQSLSTYR 60
      2  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      3  76  MYDNLQGSBGRYLYEMLGTGTPQKVOILLVDGSSNFVAGAPHSDIDYFSESSSTYH 135
      4  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      5  QY      61  DLKRGVYVYVYTGKMGEGELGDTLVSPHGPNTVAMNATIAIESPKFVNGSMGGIIGL 120
      6  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      7  DB      136  SKGFVDYVYKTIQDGMTGTFGEVDELVTIPKFNSSFLVNIATITIESEFFLPGIKMGIIIGL 195
      8  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      9  QY      121  AYAELIARPDSDLIEPFDSLKVQTHVNLFTSLDLCGAGFPLNOSSEVLASVGSMIIGIDH 180
      10  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      11  DB      196  AYAALAKPSSLETFEFDLSVAQAKIPDIEFSMGCGGLPVAS--GTNGSLVYLGIEP 252
      12  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      13  QY      181  STYTSLSMTYPIRRREMYEVIIIVREINGQDLKMDCKEYNIDKSLYDSTTLRLPKKYF 240
      14  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      15  DB      253  STYKSDIWTPTPKREMYVQIETLKLEIGQMLNLDQCREYNADKAIVDSGTTLLRPQKYF 312
      16  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      17  QY      241  EAAVESIKAASTEEFPPGFMEGLDVCQAQGTTPMNIPPEYSLVIMGEVNTQSFRTTL 300
      18  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      19  DB      313  DAVEVEVARTSLIIEPFSDFWGAQLACTNSETPTWATYFPAKSIYLRDENASRSFRITTL 372
      20  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      21  QY      301  PQQYLRPEVEDATSDDCCKFAISQSSSTGTWGVAVIMGEFVYVEDPARRRRIGFAVSAC 358
      22  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      23  DB      373  POLYIOPMVGAGFN-ECYRFGISSSTNALVIGATVMEGFVYVEDPARRRIGFAVSPC 429
      24  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-09-912-484-2
; Sequence 2, Application US/09912484
; Patent No. 6358725
; GENERAL INFORMATION:
; APPLICANT: Christie, Gary
; APPLICANT: Li, Xiaotong
; APPLICANT: Powell, David J.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Mouse Aspartic Secretase-1 (masp1)
; FILE REFERENCE: GP-70663-D1
; CURRENT APPLICATION NUMBER: US/09/912,484
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/166,974
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: 09/717,432
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 514
; TYPE: PRT
; ORGANISM: MUS MUSCULUS
; US-09-912-484-2

```

15.7%; Score 298.5; DB 1; Length 396;

TOPOLOGY: linear

IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 181994  
US-09-032-523-9

Query Match 15.7% Score 298.5; DB 4; Length 396;  
Best Local Similarity 26.7%; Pred. NO. 2e-22;  
Matches 96; Conservative 61; Mismatches 141; Indels 61; Gaps 14;

QY 13 YVEVETVSPPTQTLNIVDTGSSNFAYGA----APHPFLRYQROLSSYYRDLRKGVY 68  
D 13 YVEVETVSPPTQTLNIVDTGSSNFAYGA----APHPFLRYQROLSSYYRDLRKGVY 68  
D 78 YFGIISIGSPPTQTLNIVDTGSSNFAYGA----APHPFLRYQROLSSYYRDLRKGVY 137  
QY 69 PYTGKMEGELGDLVSIPIGPNVTANIAITSDKFFINSNMGGIIGLAEIARP 128  
D 138 QYTGSLGIGIADQVSV-EGLFVVGQGFESVTEPGQITVD-AEDDGLIGLTPSLA-- 193  
QY 129 DDLSEPFDSLVQTHVNLPSIQLCGAGPPLNOSSEVLASVGSMTIGIDHSLYTGSLM 188  
D 194 VCGTTPVFDNMMAO-----NLVDLPMSVYSSNPE--GGAGSELIIFGQYDHSHPGSLN 246  
QY 189 YTPIRREMYEVLIVRVEINGODLKMCKEYNDKSIYDSGTNLRPKKVEFAAKSK 248  
D 247 WVPYTKQAYQIALDNQVGG--TYMFCSE--GCQAIYDITSLITGSPDKIKOLONALG 302  
QY 249 AASSTEFEPGFWLGEOLVCMQAGTTPWNIPEVVISLYLMEGVNOSFRITILPOQYLREV 308  
D 303 AAP-----VGEYAVE-----CANLVMPDVTEFTING-----VPYTLSPYAV--TL 341  
QY 309 EDVATSDODCYKFAISQSTG-----TWGAVIMEGFEVVDRAKRGFA 354  
D 342 LDEVDMQFC-----SSGFQGLDIHPAGFLMILGDVFTROFYSVDRGNRRVGLA 392

## RESULT 9

US-08-208-007A-12  
Sequence 12, Application US/08208007A  
Patent No. 5501969  
GENERAL INFORMATION:  
APPLICANT: HASTINGS, ET AL.  
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER PARK ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/208, 007A  
FILING DATE: March 8, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: No. 5501969e  
FILING DATE: No. 5501969e  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 AMINO ACIDS  
TYPE: AMINO ACID

STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-208-007A-12

Query Match 15.5% Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. NO. 4.4e-22;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;

QY 13 YVEVETVSPPTQTLNIVDTGSSNFAYGAAPHPFL-----HRYQROLSSYYRDLRKGV 66  
D 79 YGEIGIGIPPCQTFVETDGTSSNLVPSIHCCKLDICWIIHKYNSDKSSYYKNGTSF 138  
QY 67 YPYTGKMEGELGDLVSIPIGPNVTANIAITSDKFFINSNMGGI 117  
D 139 DIHYSGSLSGYLSODTVPSCQASASALAGVYKQVGEANTKQGIFFIAKPDGI 198  
QY 118 LGLAYAEIARPDSDLEPFDSLVQTHV-NPLPSIQLCGAGPPLNOSSEVLASVGSMTIG 176  
D 199 LGMATPRIS--VNNVLPVFDNLMOQKLVQNIIFSEYL-----SRDPDAQPGGELMLG 248  
QY 177 GIDHSLYTGSLMYTPIREMYEVLIVRVEI-NGODLKMCKEYNDKSIYDSGTNLR 235  
D 249 GDSKYYKGSISYLNVTBKAYQWHLQVEVASGLTL---CKE--GCBAYDITSLMGV 303  
QY 236 PKYVEFAVKSIRKAASSTEFEPGFWLGEOLV-CWQAGTTPWNIPEVVISLYLMEGVNOS 294  
D 304 PVDEYRELQKAIQAVPLIQ-----GEYMIPEKYST-----LPAITLKLGG---KG 346  
QY 295 FRITILPOQYLREPVEDVATSDODCYKFAISQ-----SSTGVGAVIMEGFEVYFEDARK 349  
D 347 YKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIPPGPGLMILGDVFTIGRYTVFDRDN 402  
QY 350 RIGFAVSA 357  
D 403 RVGFARAA 410

## RESULT 10

US-08-974-691-4  
Sequence 4, Application US/08974691  
Patent No. 6225103  
GENERAL INFORMATION:  
APPLICANT: Keolsch, Gerald  
APPLICANT: Lin, Xinli  
APPLICANT: Tang, Jordan  
TITLE OF INVENTION: Cloning and Characterization of Napsin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 W. Peachtree  
STREET: St.  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,691  
FILING DATE: 20-NOV-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/031,196  
FILING DATE: 20-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,126  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.



Best Local Similarity 25.3%; Pred. No. 8,1e-20;  
Matches 92; Conservative 59; Mismatches 144; Indels 69; Gaps 11;

QY 13 YVEMTVGSPQTLNLTVDGSSNFAVGAAP-----HPLHRYQROSLSTYRDLRKGYV 68  
Db 96 YFTEITLGSPPQSFKYLIDTGSNLMWPFAEGSLACFLHRTKYDHEASTYKANGSEPAI 155  
QY 69 PYQGWKEGELGTDLVSIPIHGNVTVRANIAITESDKFFINGSNMEGILGLAETIARP 128  
Db 156 QVSGSLEGVSRDLTLTI--GDLVIPDQFAEATSEPLAFAGFKFGDILGLADYSIS-V 212  
QY 129 DDLSEFFDSLVKQTHVNLFSLOLCGAGFPLNOSSEVLASVGSMTIIGIDHSLYTGSIW 188  
Db 213 NRIVPPVYNAIKNNKLDDPVFA-----FYLGDSD-KSEGGGASFGGIDEKTYTEIT 264  
QY 189 YPIRREMYEYIIVREINQDLKMDCKEYNDKSIYDSGTNLRPKKVFEAIVSIK 248  
Db 265 WLPVRRKAWE-----VFEGIGLGEYATLEGHGAIDTGSTLALPSGLAETIARNAIG 319  
QY 249 AASSTKFPDGFNLGEOLVCMQAGTTPWNIFFPVISLYLMGEVTVNOSFRIT----- 298  
Db 320 AKKG-----WSQYSVDCESRDS-----LPDLTLNFG-----YNFTIAYDYTLEVSG 363  
QY 299 -----ILPQOYLPRVEDVATSDQCYKFAISOSSTGVMGAVIMEGYVYVDRARRKRGF 353  
Db 364 SCISAFTEPMDPEPVGPLA-----IDGDAFLRKRYSTYIDIGHDAVGL 405  
QY 354 AVSA 357  
Db 406 AKAA 409

RESULT 13

US-08-360-673-6  
Sequence 6, Application US/08360673  
Patent No. 5679544

GENERAL INFORMATION:

APPLICANT: Fleer, Reinhard  
APPLICANT: Fournier, Alain  
APPLICANT: Yeh, Patrice  
TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19002

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/360, 673  
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR93/00623  
FILING DATE: 23-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/07785  
FILING DATE: 25-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Julie K.  
REGISTRATION NUMBER: 38,619  
REFERENCE/DOCKET NUMBER: ST92040-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)454-3839  
TELEFAX: (610)454-3808

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 409 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-360-673-6

Query Match 14.3%; Score 273.5; DB 1; Length 409;  
Best Local Similarity 25.3%; Pred. No. 8,1e-20;  
Matches 92; Conservative 59; Mismatches 144; Indels 69; Gaps 11;

QY 13 YVEMTVGSPQTLNLTVDGSSNFAVGAAP-----HPLHRYQROSLSTYRDLRKGYV 68  
Db 96 YFTEITLGSPPQSFKYLIDTGSNLMWPFAEGSLACFLHRTKYDHEASTYKANGSEPAI 155  
QY 69 PYQGWKEGELGTDLVSIPIHGNVTVRANIAITESDKFFINGSNMEGILGLAETIARP 128  
Db 156 QVSGSLEGVSRDLTLTI--GDLVIPDQFAEATSEPLAFAGFKFGDILGLADYSIS-V 212  
QY 129 DDLSEFFDSLVKQTHVNLFSLOLCGAGFPLNOSSEVLASVGSMTIIGIDHSLYTGSIW 188  
Db 213 NRIVPPVYNAIKNNKLDDPVFA-----FYLGDSD-KSEGGGASFGGIDEKTYTEIT 264  
QY 189 YPIRREMYEYIIVREINQDLKMDCKEYNDKSIYDSGTNLRPKKVFEAIVSIK 248  
Db 265 WLPVRRKAWE-----VFEGIGLGEYATLEGHGAIDTGSTLALPSGLAETIARNAIG 319  
QY 249 AASSTKFPDGFNLGEOLVCMQAGTTPWNIFFPVISLYLMGEVTVNOSFRIT----- 298  
Db 320 AKKG-----WSQYSVDCESRDS-----LPDLTLNFG-----YNFTIAYDYTLEVSG 363  
QY 299 -----ILPQOYLPRVEDVATSDQCYKFAISOSSTGVMGAVIMEGYVYVDRARRKRGF 353  
Db 364 SCISAFTEPMDPEPVGPLA-----IDGDAFLRKRYSTYIDIGHDAVGL 405  
QY 354 AVSA 357  
Db 406 AKAA 409

RESULT 14

US-08-846-021A-8  
Sequence 8, Application US/08846021A  
Patent No. 5948682

GENERAL INFORMATION:

APPLICANT: Moloney, Maurice M.  
TITLE OF INVENTION: Preparation of Heterologous Proteins on  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/846,021A  
FILING DATE: April 25, 1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Michelle  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 9369-039  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398

INFORMATION FOR SEQ ID NO: 8:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 427 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-021A-8

```

```

Query Match          14.3%; Score 273.5; DB 2; Length 427;
Best Local Similarity 26.4%; Pred. No. 8,6e-20;
Matches 94; Conservative 68; Mismatches 129; Indels 65; Gaps 15;

```

```

QY 13 YVEMTVGSPQDTLNIIVDTGSSNFAYGAAPHPFL-----HRYQROLSTYRDLKRG 65
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 120 YFGKIYLTGPQETVLEFDTSDDPWV---PSIYCKSNACKNHQRFDPKRSSTFQNLGKP 176
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 66 YVPTQKWEGLGELDTLVSTPHGPNVTVRANIAITESDKFINGSNWGGIIGLAYAEI 125
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 177 LSIHGTSGMOGIIIGYDVTVSN--IVDIQOTVGLSTQEPEDVTFYAEFDIIGMAYPSL 234
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 126 ARPDLSLEPFPDSLVKQTHV--PNLFSLQCGAFPLNOSVYLVASVGSMTIGIDHSLYT 184
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 235 A--SESTIPVFDMNMRHLVAQDLFSVMDRNG-----QESMLT-----LGAIIDPSYTT 281
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 185 GSLWYTPPIREMYEYEVIIIVREINGQDLKMD--CKEYVYDKSIYDSGTTNLRPKVFEA 242
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 282 GSLHWVPVTVQOYQWQFTVDSTISGVYVACEGCG-----QALIDTGSKLVGSPSDDL 335
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 243 AKSIKAASSTKFPDGFVLGE--QLVCWQAGTTPMNIFFPVISLYLMGEVYNOSFRITLIP 301
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 336 IQAIGATQNO-----YGFEDIDCDNLSTYMTVF-----EINGKMYPLT--P 376
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 302 QQYLRVEDVATSQDDCY---KFAISQSGTGYMGAVIMEGFYVEDRARRKIGFAV 354
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 377 SAT-----TSQDQGFCTSGFQSENHSOKMLGDVFIREFYVDFRANNLVGIA 424
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 15
US-08-088-633-2
; Sequence 2, Application US/08088633
; Patent No. 5324660

```

```

; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; TITLE OF INVENTION: Genes which influence Pichia proteolytic
; TITLE OF INVENTION: Activity, and uses therefor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 South LaSalle Street, Suite 900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/088,633
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: 50848
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)552-1311
; TELEFAX: (619)552-0095
; FILEX: 20 6566 PATLAW CGO

```

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-088-633-2

```

```

Query Match          14.2%; Score 270; DB 1; Length 410;
Best Local Similarity 25.7%; Pred. No. 1,9e-19;
Matches 91; Conservative 51; Mismatches 162; Indels 50; Gaps 10;

```

```

QY 13 YVEMTVGSPQDTLNIIVDTGSSNFAYGA---APHPILHYQROLSTYRDLKRGVYV 68
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 96 YTFEVSIGTPPSFKVILDTGSSNLMVPSKDCGSLACFLAKYHDDSSYIKKNGSSFEI 155
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 69 YPTQKWEGLGELDTLVSTPHGPNVTVRANIAITESDKFINGSNWGGIIGLAYAEIAR 128
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 156 RYSGSGMEGYVSQDVLOI--GDLTIPIKVDFAEATSEPGIARAFGKFDGIIIGLAY----- 207
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 129 DQSLLEFPDLSLVKQTHVPLF--SLQLCGA---GFPLNOSVYLVASVGSMTIGIDHSL 182
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 208 -----DSISVNRKIIVPIYKALBELDLDEPKFAFYIGTDKDESDGGLATFGGVQDKSK 259
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 183 YTGSLWYTPPIREMYEYEVIIIVREINGQDLKMDCKEYVYDKSIYDSGTTNLRPKVFEA 242
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 260 YEKRTIWLPIVRKRAYE-----VSFDGVLGSEYVLELQKTAIDTGSILALPSGLAEI 314
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 243 AVSIKAASSTKFPDGFVLGEQLVCWQAGTTPMNIFFPVISLYLMGEVYNOSFRITLIP 302
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 315 LMEIGATFG-----WQGYAVVDCDTRDS---LPDLTLTFAG-----YNFTIIPY 356
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 303 QYLRVEDVATSQDDCYKFAISQSGTGYMGAVIMEGFYVEDRARRKIGFAV 356
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 357 DYTLEVSGSCISAFTPMDFP--EPIGPLAITGDSFLKRYYSVYDLGDAVGLAKS 409
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Search completed: October 30, 2002, 12:32:32
Job time : 54.4447 secs

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:10 ; Search time 31.0442 Seconds

(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-75

Perfect score: 1907

Sequence: 1 MVDLKRKSGGGYVENTVG.....VFDRARRKRGFAVSACHVH 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProtL40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1907	100.0	501	1	BACE_HUMAN
2	1896	99.4	501	1	BACE_MOUSE
3	1890	99.1	501	1	BACE_RAT
4	1106	58.0	518	1	BACE2_HUMAN
5	327	17.1	324	1	PEP1_GADMO
6	306.5	16.1	390	1	CATD_BOVIN
7	306	16.0	387	1	PEP1_RABIT
8	302	15.8	367	1	PEP1_CHICK
9	302	15.8	388	1	PEP4_CHICK
10	298.5	15.7	383	1	PEP4_MACFU
11	298.5	15.7	396	1	PEPE_CHICK
12	296	15.5	387	1	CATE_HUMAN
13	296	15.5	391	1	PEP2_RABIT
14	295.5	15.5	412	1	CATE_CAVPO
15	295	15.5	387	1	CATD_HUMAN
16	287	15.0	407	1	PEP4_RABIT
17	286	15.0	387	1	CATD_RAT
18	285.5	15.0	387	1	PEP3_RABIT
19	285	14.9	388	1	CATE_RAT
20	285	14.9	398	1	PEP2_MACFU
21	283	14.8	388	1	CATD_CHICK
22	281.5	14.8	381	1	PEP4_MACMU
23	281	14.7	388	1	CHYM_SHEEP
24	279.5	14.7	410	1	PEP1_MACFU
25	278.5	14.6	386	1	PEP4_MOUSE
26	278	14.6	387	1	PEP4_PIG
27	278	14.6	388	1	PEP4_CALJA
28	277.5	14.6	396	1	PEP4_HUMAN
29	277.5	14.6	397	1	CATE_RABIT
30	274	14.4	419	1	CATE_MOUSE
31	273.5	14.3	381	1	CARV_BOVIN
32	273	14.3	388	1	CHYM_BOVIN
33	270.5	14.2	377	1	PEP3_RABIT
					PEP3_MACFU

34	270	14.2	376	1	PAG2_BOVIN
35	267.5	14.0	381	1	CHYM_CALJA
36	266	13.9	405	1	CARP_YEAST
37	264	13.8	396	1	CARP_NEUCR
38	263.5	13.8	388	1	PEP4_CALJA
39	262.5	13.8	394	1	PEP4_CAVPO
40	261.5	13.7	387	1	ASPP_HUMAN
41	261.5	13.7	388	1	PEP4_HUMAN
42	261	13.7	388	1	PAG_HORSE
43	260	13.6	345	1	CATD_PIG
44	259	13.6	365	1	CATD_SHEEP
45	250.5	13.1	496	1	ASPP_ORYSA

## ALIGNMENTS

RESULT 1	ID	Sequence	Standard	PRT	501 AA
BACE_HUMAN	1	1907	100.0	501	1
BACE_HUMAN	2	1896	99.4	501	1
BACE_HUMAN	3	1890	99.1	501	1
BACE_HUMAN	4	1106	58.0	518	1
BACE_HUMAN	5	327	17.1	324	1
BACE_HUMAN	6	306.5	16.1	390	1
BACE_HUMAN	7	306	16.0	387	1
BACE_HUMAN	8	302	15.8	367	1
BACE_HUMAN	9	302	15.8	388	1
BACE_HUMAN	10	298.5	15.7	383	1
BACE_HUMAN	11	298.5	15.7	396	1
BACE_HUMAN	12	296	15.5	387	1
BACE_HUMAN	13	296	15.5	391	1
BACE_HUMAN	14	295.5	15.5	412	1
BACE_HUMAN	15	295	15.5	387	1
BACE_HUMAN	16	287	15.0	407	1
BACE_HUMAN	17	286	15.0	387	1
BACE_HUMAN	18	285.5	15.0	387	1
BACE_HUMAN	19	285	14.9	388	1
BACE_HUMAN	20	285	14.9	398	1
BACE_HUMAN	21	283	14.8	388	1
BACE_HUMAN	22	281.5	14.8	381	1
BACE_HUMAN	23	281	14.7	388	1
BACE_HUMAN	24	279.5	14.7	410	1
BACE_HUMAN	25	278.5	14.6	386	1
BACE_HUMAN	26	278	14.6	387	1
BACE_HUMAN	27	278	14.6	388	1
BACE_HUMAN	28	277.5	14.6	396	1
BACE_HUMAN	29	277.5	14.6	397	1
BACE_HUMAN	30	274	14.4	419	1
BACE_HUMAN	31	273.5	14.3	381	1
BACE_HUMAN	32	273	14.3	388	1
BACE_HUMAN	33	270.5	14.2	377	1

"Identification of a novel aspartic proteinase (Asp 2) as beta-secretase.";  
 Mol. Cell. Neurosci. 14:419-427(1999).  
 [5]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Brain, and Pancreas;  
 RA Michel B., De Pietri Tonelli D., Zaccchetti D., Keller P.;  
 RT "New beta-site APP cleaving enzyme isoform (BACE-1b) obtained from  
 human brain and pancreas.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.  
 RA MEDLINE=20144060; PubMed=10677483;  
 RX Lin X., Koelsch G., Wu S., Downs D., Dashed A., Tang J.;  
 RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of  
 beta-amyloid precursor protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
 CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: A/BACE-1a (shown here) and  
 CC B/BACE-1b; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EURAROTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF190725; AAF04142.1; -  
 CC EMBL: AF201468; AAF18982.1; -  
 CC EMBL: AF200343; AAF17079.1; -  
 CC EMBL: AF204943; AAF26367.1; -  
 CC EMBL: AF338816; AAK38374.1; -  
 CC EMBL: AF200193; AAF13715.1; -  
 CC HSP: P32329; IYPS.  
 CC MIM: 604252; -  
 CC DR InterPro: IPR001969; Asp-protease.  
 CC DR InterPro: IPR001461; Pepsin.  
 CC DR Pfam: PF00026; asp; 3.  
 CC DR PRINTS: PR00792; PEPsin.  
 CC DR PROSITE: PS00141; Asp-PROTEASE; 1.  
 CC KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
 CC Signal; Alternative splicing.  
 CC FT SIGNAL 1 21  
 CC FT PROPEP 22 45  
 CC FT CHAIN 46 501  
 CC FT DOMAIN 22 457  
 CC FT TRANSMEM 458 478  
 CC FT DOMAIN 479 501  
 CC FT ACT\_SITE 93 93  
 CC FT ACT\_SITE 289 289  
 CC FT CARBOHYD 153 153  
 CC FT CARBOHYD 172 172  
 CC FT CARBOHYD 223 223  
 CC FT CARBOHYD 354 354  
 CC FT VARSLIC 190 214  
 CC SEQUENCE 501 AA: 55763 MW: 37064824ACEF05 CRC64:  
 BETA-SECRETASE.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 BY SIMILARITY.  
 BY SIMILARITY.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 MISSING (IN ISOFORM B).  
 MISSING (IN ISOFORM B).

Query Match 100.0%; Score 1907; DB 1; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 6e-153;  
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MVDLNRKSGGYYVEMTVGSPQTLNIIIVDTGSSNFVAGADPHPLHRYQRQLSSTYR 60

Db 63 MVDLNRKSGGYYVEMTVGSPQTLNIIIVDTGSSNFVAGADPHPLHRYQRQLSSTYR 122  
 Oy 61 DLKRGVYPTTQGGKEGELDTLVSIPHGPNVYVRANIAITSDFFINGSMWEGILGL 120  
 Db 123 DLKRGVYPTTQGGKEGELDTLVSIPHGPNVYVRANIAITSDFFINGSMWEGILGL 182  
 Oy 121 AYAEIARPDLSLEFFPDLSVKQTHVPLFSLQAGAPFNPOSEVYLASVGGSMITIGIDH 180  
 Db 183 AYAEIARPDLSLEFFPDLSVKQTHVPLFSLQAGAPFNPOSEVYLASVGGSMITIGIDH 242  
 Oy 181 SLYTGSLMYPPIREMYEYIIVAVEINSGDLKDKCEYVNDKSYNSGTNRLPKVFE 240  
 Db 243 SLYTGSLMYPPIREMYEYIIVAVEINSGDLKDKCEYVNDKSYNSGTNRLPKVFE 302  
 Oy 241 EAAKSTKAASSTKFPDGFGLGOLCOWAGTTPMNIFFPISLYLMGEVYNOSFRITIL 300  
 Db 303 EAAKSTKAASSTKFPDGFGLGOLCOWAGTTPMNIFFPISLYLMGEVYNOSFRITIL 362  
 Oy 301 PQYLRLPEVDVATSDQDCYKFAISQSSGTGVMGAVIMEGFYVVDRAKRRIGFAVSACHV 360  
 Db 363 PQYLRLPEVDVATSDQDCYKFAISQSSGTGVMGAVIMEGFYVVDRAKRRIGFAVSACHV 422  
 Oy 361 H 361  
 Db 423 H 423  
 RESULT 2  
 BACE-MOUSE STANDARD: PRT: 501 AA.  
 AC P56818;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
 DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
 DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)  
 DE (Memapsin-2).  
 DE BACE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20002972; PubMed=10531052;  
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
 RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
 RA Treanor J., Rogers G., Citron M.;  
 RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
 RT the transmembrane aspartic protease BACE.";  
 RL Science 286:735-741(1999).  
 RN [2]  
 RP REVISIONS TO 6 AND 81-87.  
 RA Bennett B.D., Vassar R., Citron M.;  
 RA submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057170; PubMed=10591213;  
 RA Yan R., Blenkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,  
 RA Braslier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
 RA Tomasselli A.G., Parodi L.A., Heinrichson R.U., Guney M.E.;  
 RA "Membrane-anchored aspartyl protease with Alzheimer's disease  
 RT beta-secretase activity.";  
 RL Nature 402:533-537(1999).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL

```

CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: BRAIN
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF190726; AAF0143.2; -
DR EMBL: AF200346; AAF17082.1; -
DR HSSP: P56272; IAMS.
DR MGD: MGI:1346542; Bace.
DR InterPro: IPR001969; Asp-protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF000026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KM Hydrolyse: Aspartyl protease; Zymogen; Transmembrane;
KM Signal.
FT SIGNAL. 1 21 POTENTIAL.
FT PROPEP 22 45 POTENTIAL.
FT CHAIN 46 501 BETA-SECRETASE.
FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 458 478 POTENTIAL.
FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 289 289 BY SIMILARITY.
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 501 AA; 55747 MW; C085A013145E474E CRC64;

Query Match 99.4%; Score 1896; DB 1; Length 501;
Best Local Similarity 98.6%; Pred. No. 5.1e-152;
Matches 356; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

RESULT 3
BACE_RAT
ID BACE_RAT STANDARD: PRT: 501 AA.
AC P56819;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)
DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl
DE protease 2) (Asp 2) (Asp2) (Membrane-associated aspartic protease 2)
DE (Memapsin-2).
DE BACE.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amaratne P., Loeflof R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE."
RL Science 286:735-741(1999).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,
CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF190727; AAF0144.1; -
DR HSSP: P32329; IYPS.
DR InterPro: IPR001969; Asp-protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF000026; asp. 3.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
KM Hydrolyse: Aspartyl protease; Zymogen; Transmembrane;
KM Signal.
FT SIGNAL. 1 21 POTENTIAL.
FT PROPEP 22 45 POTENTIAL.
FT CHAIN 46 501 BETA-SECRETASE.
FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 458 478 POTENTIAL.
FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 289 289 BY SIMILARITY.
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 501 AA; 55806 MW; 24B445BC8B87DE3 CRC64;

Query Match 99.1%; Score 1890; DB 1; Length 501;
Best Local Similarity 98.3%; Pred. No. 1.6e-151;
Matches 355; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 MYDNLRGSGGQYVYEMVSGSPOTLNIIVDTGSSNFVAGAPHPFLHRYROLSTYR 60  
 DB 63 MYDNLRGSGGQYVYEMVSGSPOTLNIIVDTGSSNFVAGAPHPFLHRYROLSTYR 122  
 OY 61 DLKRGVYVPTOGKWEGLGTDLVSIPIHGPVNTVRANIAAITESDKEFLINGSNMEGIIGL 120  
 DB 123 DLKRGVYVPTOGKWEGLGTDLVSIPIHGPVNTVRANIAAITESDKEFLINGSNMEGIIGL 182  
 OY 121 AVALIARPDSDLEPPFDLSLVKOTHPNLFSLDLCGAPPLNDSVYLAVSGSMITIGIDH 180  
 DB 183 AVALIARPDSDLEPPFDLSLVKOTHPNLFSLDLCGAPPLNDSVYLAVSGSMITIGIDH 242  
 OY 181 SLYTGSLWTPTRRREYVYVILVIREINQODLKMCKEYNYKSYDSCGTMRLPKKYF 240  
 DB 243 SLYTGSLWTPTRRREYVYVILVIREINQODLKMCKEYNYKSYDSCGTMRLPKKYF 302  
 OY 241 EAAVSIKASSTKPEFDPGFWLGEOLVCMQAGTPWNIFPVISLYLMGEVNTQSPRITTL 300  
 DB 303 EAAVSIKASSTKPEFDPGFWLGEOLVCMQAGTPWNIFPVISLYLMGEVNTQSPRITTL 362  
 OY 301 PQOYLRPVEDVATSDDDCKKFAISOSTGTVMGAVIMEGFYVPEPRARRIGFAVSACHV 360  
 DB 363 PQOYLRPVEDVATSDDDCKKFAISOSTGTVMGAVIMEGFYVPEPRARRIGFAVSACHV 422  
 OY 361 H 361  
 DB 423 H 423

## RESULT 4

BAE2\_HUMAN STANDARD: PRT; 518 AA.

ID BAE2\_HUMAN OGUJTB6;  
 AC OGY520: 2000 (rel. 39, Created)  
 DT 30-MAY-2000 (rel. 39, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)  
 DE Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving enzyme 2) (Aspartyl protease 1) (Asp 1) (Asp1) (Membrane-associated aspartic protease 1) (Memapsin-1).  
 GN BACE2 OR ASP21.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057170; PubMed=10591213;  
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M., Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B., Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;  
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity."; Nature 402:533-537(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RA Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U., Giese K.;  
 RT "Identification of a novel aspartic-like protease differentially expressed in human breast cancer cell lines."; Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RP Accorino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;  
 RT "Cloning of a gene from chromosome 21 Down region encoding a potential transmembrane aspartyl protease."; Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN SEQUENCE FROM N.A.  
 RP Solans A., Estivill X., de la Luna S.;  
 RT "Cloning of a novel mammalian aspartyl protease."; Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RL [5]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20030166; PubMed=10561122;  
 RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D., Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M., Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;  
 RT "Identification of a novel aspartic proteinase (asp 2) as beta-secretase."; Mol. Cell. Neurosci. 14:419-427(1999).  
 RL [6]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=20144060; PubMed=10677483;  
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;  
 RT "Human aspartic protease mepsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein."; Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).  
 RL [7]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=20289799; PubMed=10830953;  
 RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P., Schafte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E., Wehrmeyer S., Borzay K., Gardiner K., Nizetic D., Francis F., Lehmann H., Reinhardt R., Yaspo M.-L.;  
 RT "The DNA sequence of human chromosome 21." Nature 405:311-319(2000).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF200342; AAF17078.1; -  
 DR EMBL: AF117892; AAD45240.1; -  
 DR EMBL: AF050171; AAD45963.1; -  
 DR EMBL: AF178532; AAF28494.1; -  
 DR EMBL: AF204944; AAF28368.1; -  
 DR EMBL: AF200192; AAF13714.1; -  
 DR EMBL: AL163284; CAB90458.1; -  
 DR EMBL: AL163285; CAB90554.1; -  
 DR HSP: P00797; 2REN.  
 DR MIM: 605668; -  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp; 3.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane; Signal.  
 KW  
 FT SIGNAL 1 20  
 FT PROPEP 21 ?  
 FT CHAIN ? 518  
 FT DOMAIN 21 473  
 FT TRANSMEM 474 494  
 FT DOMAIN 495 518  
 FT ACT\_SITE 110 110  
 FT ACT\_SITE 303 303  
 FT CARBOHYD 170 170  
 FT CARBOHYD 366 366  
 FT CONFLICT 36 A -> T (IN REF. 6).  
 SQ SEQUENCE 518 AA; 56180 MW; 2E903150823760D3 CRC64;

```

Query Match          58.0%; Score 1106; DB 1; Length 518;
Best Local Similarity 56.4%; Pred. No. 1.6e-85;
Matches 202; Conservative 60; Mismatches 92; Indels 4; Gaps 2;

OY 1 MVDNLRGKSGGYVEMVYSPROTLLIVDTGSSNFAGNAPHPFLHRYRQISSTYR 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 80 MVDNLQGDGSRGYLEMLGTPQKQLIVDTGSSNFAGNAPHPFLHRYRQISSTYR 139
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 61 DLKRGVYVYTGKWEGLGTDLVSIHPGPNVTVRANIAIESKFFLNGSMWGIIGL 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 140 SKGEVTVYKVTGSGWTFGEGLDVLITPKGFNTSLFNATIFESENFLPGIKMNGIIGL 199
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 121 AAELARPDSDLEPPFDSLVKOTHPNLFSIQLCGAGPPLNQSSEVLASVSGSMITGIDH 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 200 AVATLAKPSSSLETFEFDLSVTCANIPNVFSMQMGAGLPLVAGS---GNGSGSLVGLIEP 256
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 181 SLYTGSLWVTPFRREMYEVLIVRVEINGODLKMCKEYNDKSLVDSGTTLRLPKRYF 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 257 SLYKGDIVTTPKREMYVQIEILKLEIGQSINLDCREVMADKALVDGFTLLRLPKRYF 316
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 241 EAAVKSINAASTEFKPDGFWLGEQLVCMQAGTTPMNIFFVLSYLMGEVTVNSFRITL 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 317 DAVBAVARASLIPEFSOGFWTGSOLACMTNSETPMWYFPAKISTYLRDENSRSRRTITL 376
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 301 PQQYLRPEVDVATSDDCYKFAISQSTGTWYGAIVMEGFYVVEDRARRKRGFAVSAC 358
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 377 POLYIOPMKGAGLNY-ECYRFGISPTNALVIGATVMEGFYVIFDRAQKRVGFAASPC 433
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
PEPL_GADMO STANDARD; PRT; 324 AA.
ID PEPL_GADMO
AC P56272;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Pepsin IIB (EC 3.4.23.-).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
RC TISSUE=Stomach;
RA Karlsten S., Hough E., Olsen R.L.;
RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
RT cod (Gadus morhua).";
RL Acta Crystallogr. D 54:32-46(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PDB: 1AM5; 24-DEC-97.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp_1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW 3D-structure.
FT ACT_SITE 32 BY SIMILARITY.
FT ACT_SITE 214 BY SIMILARITY.
FT DISULFID 45 BY SIMILARITY.
FT DISULFID 206 BY SIMILARITY.
FT DISULFID 209 BY SIMILARITY.
FT DISULFID 247 BY SIMILARITY.
FT DISULFID 280 BY SIMILARITY.
SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match          17.1%; Score 327; DB 1; Length 324;
Best Local Similarity 27.9%; Pred. No. 3.1e-20;
Matches 104; Conservative 66; Mismatches 137; Indels 66; Gaps 15;

OY 1 MVDNLRGKSGGYVEMVYSPROTLLIVDTGSSNFAGNAPHPFLHRYRQISSTYR 56

```

```

DB 2 VTEQMKNEADTEYGVVISIGTPPESFKVIFDTGSSNLMWSSSHCSAQACSNNHKKFRRQS 61
OY 57 STYRLRGVGVVPTQKWEGLGTDLVSIHPG--PNVTVRANIAATIESKFFLNGSM 114
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 62 STYETGKTVLDLTGTGGMRGILGODIVYSGGSDPNDELG---ESQEPPEPQA-AAPE 117
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 115 EGIIGLAAELARPDSDLEPPFDSLVKOTHP-NPLFSIQLCGAGPPLNQSSEVLASVSGSM 173
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 118 DGIIGLALPSTAA--GAVPYFDNNGSGLYEKDLFSTYLSGGC--ANGSEVM----- 166
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 174 IIGGIDHSLYGLSWYPIRREMYEVLIVRVEINGODLKMDCREYNDKSLVDSGTTL 232
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 167 LGGVNDNSHYGSIHWIPVTAKEYQWVALDGIYVNGTAAACEG---QAIYDTGTSK 219
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 233 LRLPKRYEAAVKSINAASTEFKPDGFWLGEQLVCMQAGTTPMNIFFVLSYLMGEVTVN 292
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 220 IYAPVSALANIMKDIGASEN-----QGEIMGN---CASVQSLPDTF-----TI 260
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 293 QSFRTIIPQQLYLRPEVDVATSDDCYKFAISQSTGT-----VMGAVIMEGFYVVF 344
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 261 NGVQPLPSPSAVIEGDAQFCTS-----GLSSGVPSNTSELMIFGDVFLNNYTYI 311
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 345 DRARRKRGFAVSA 357
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 312 DRTNKKVGFADAA 324
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 6
CATD_BOVIN STANDARD; PRT; 390 AA.
ID CATD_BOVIN
AC P80209; O9TS27;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cathepsin D precursor (EC 3.4.23.5).
GN CTSD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-48.
RC TISSUE=Milk;
RX MEDLINE=93202276; PubMed=8454061;
RA Larsen L.B., Bolisen A., Petersen T.E.;
RT "Procathepsin D cannot autoactivate to cathepsin D at acid pH.";
RL FEBS Lett. 319:54-58(1993).
RN [2]
RP SEQUENCE OF 45-390, AND X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RC TISSUE=Liver;
RX MEDLINE=93223670; PubMed=8467789;
RA Metcalf P., Fusek M.;
RT "Two crystal structures for cathepsin D: the lysosomal targeting
RT signal and active site.";
RL EMBO J. 12:1293-1302(1993).
CC -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
CC PROTEIN BREAKDOWN.
CC -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
CC that of pepsin A. Does not cleave the 4-Gln-His-5 bond in B
CC chain of insulin.
CC -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR; S3283; S32383.
DR PIR; S37419; S37419.
DR HSSP; P07339; ILVB.
DR MEROPS; A01.009; -.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam; PF00026; asp_1.

```

PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE: 2.  
 KW Hydrolyase; Aspartyl protease; Glycoprotein; Lysosome; Zymogen.  
 FT PROPEP 1 44 ACTIVATION PEPTIDE.  
 FT CHAIN 45 390 CATHEPSIN D.  
 FT ACT\_SITE 77 77  
 FT ACT\_SITE 77 77  
 FT ACT\_SITE 273 273  
 FT ACT\_SITE 71 140  
 FT DISULFID 90 97  
 FT DISULFID 264 268  
 FT DISULFID 307 344  
 FT DISULFID 114 114  
 FT CARBOHYD 241 241  
 FT SEQUENCE 390 AA; 4248 MW; 5B38A1C33C48D35 CRC64;  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 390 AA; 4248 MW; 5B38A1C33C48D35 CRC64;  
 Query Match 16.1%; Score 306.5; DB 1; Length 390;  
 Best Local Similarity 28.1%; Pred. No. 2.1e-18;  
 Matches 107; Conservative 69; Mismatches 118; Indels 87; Gaps 19;  
 Oy 13 YVEMTVGSPQTLNIIIVDGGSNFAVGAPHPFL-----HRYGKQLSSTYRDLRKGV 66  
 Db 59 YVGGIGTPOCTTVFVFDGSSANLWVPSHCKLIDLACWTHRKYNDSKSTY--VANGT 116  
 Oy 67 Y--VPYTGKMEGELGDLVSIPIHGPN-----VTVRANI--AAITSDKFFINGSMNG 116  
 Db 117 TFDIHGSGSLGSLGSDTVSPCNSSSPGVTYVQRTGGAIKQGVYFI-AAKEDG 175  
 Oy 117 ILGLAVAIAPDPSLEFPDLSLKQTHV-PNLSFLOLCAGFPINLOSEVLASVGGSMII 175  
 Db 176 ILGMAVAPRIS--VNVNLEVFENLMQOKLVDRKVS-----FELNR--DPAQPGGELML 225  
 Oy 176 GGDHSITGSLWTPPIREWEYEVIIIVVEINGODLMDCKEYNDKSIYDSGTNRL 235  
 Db 226 GGTDSKYTRGSLFNFNTRQAWQIHMDDLV-GSSLVV-CK--GGCAIYDTGSLIVG 281  
 Oy 236 PKKVEAAVKSIRKAASTKRPDGFGLGEOLV-CWQAGTTPWNIIPVLSLIMGVTVMS 294  
 Db 282 PVEEVRLEOKAIGAVPLIQ-----GEYMIPEKRVSS--LPEYTVKLG----- 322  
 Oy 295 FRITTLPOQLRPVEDVATSQDDCYKFAISOSSTGT-----VMGAVI 336  
 Db 333 -----KDYALSPED-YALKVQAETTVCLSGFGMDIPPGGPIMLIGDYF 367  
 Oy 337 MEGFYVVEFDRARRIGFAVSA 357  
 Db 368 IGRYTVFDRQNRVGLAEAA 388  
 RESULT 7  
 PEP1\_RABIT STANDARD; PRT; 387 AA.  
 ID PEP1\_RABIT  
 AC P28712:  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Pepsin II-1 precursor (EC 3.4.23.1) (Pepsin A).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91009127; PubMed-2129536;  
 RA Kagiyama T., Tanabe K., Koiwai O.;  
 RT "Structure and development of rabbit pepsinogens. Stage-specific  
 RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and  
 RT gene expression during development."  
 RL J. Biol. Chem. 265:17031-17038(1990).  
 CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-  
 CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE

THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.  
 CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPsin GENES IS REGULATED BY  
 CC HORMONES AND RELATED SUBSTANCES.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 DR PTR, B38302; B38302.  
 DR HSP; P00791; 1PSA.  
 DR MEROPS; A01.001; -;  
 DR InterPro; IPR001969; Asp-protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp. 1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE: 2.  
 KW Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Signal; Phosphorylation; Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 59 ACTIVATION PEPTIDE.  
 FT CHAIN 60 387 PEPsin II-1.  
 FT MOD\_RES 129 129 PHOSPHORYLATION (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 276 276 BY SIMILARITY.  
 FT DISULFID 106 111 BY SIMILARITY.  
 FT DISULFID 267 271 BY SIMILARITY.  
 FT DISULFID 310 343 BY SIMILARITY.  
 SO SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;  
 Query Match 16.0%; Score 306; DB 1; Length 387;  
 Best Local Similarity 27.1%; Pred. No. 2.3e-18;  
 Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;  
 Oy 13 YVEMTVGSPQTLNIIIVDGGSNFAVG---AAPHPILHRYGKQLSSTYRDLRKGVV 68  
 Db 75 YFGTISGTPQERTVIFDGGSNLWVPSYICSLACFLHFRPDDSTFGAATSELSI 134  
 Oy 69 PYTGKMEGELGDLVSIPIHGPNTVRANIATIESD---KEFINGSMGICLGLAEI 125  
 Db 135 TYGGSMTGILGYFVAKV---GNIEDTQIPLGSKTEGTGTFV--APFDILGLAVPSI 189  
 Oy 126 ARPDPSLEFPDLSLKQTHV-PNLSFLOLCAGFPINLOSEVLASVGGSMIIGIDHSY 184  
 Db 190 SASDAT--PVPDMMNMEGLVSEDLFVSYSNG-----EKGSVMVGGIDSSYIT 237  
 Oy 185 GSWLTPPIREWEYEVIIIVVEINGODLKM--DCKEYNDKSIYDSGTNRLPKVFEA 242  
 Db 238 GSLMNVVPSHSCYQIOTMDSTINGETIACADSC-----QAVYDTGSLAGFSAISK 291  
 Oy 243 AVKSIRKAASTKRPDGFGLGEOLV-CWQAGTTPWNIIPVLSLIMGVTVMSRITLP 301  
 Db 292 IQSYIGASKNL-----LGENITCSAIDSLPDIVF-----TINN 325  
 Oy 302 QOYLRPVED-VATSQDDC---YKFAISOSSTGT---VMGAVIMEGFYVVEFDRARRIGFAV 355  
 Db 326 VQYPPASAYITLKEDDCLSGFGKMLDTSTGELMILDDVFLRQYTFYFDRANNGVGLAA 385  
 Oy 356 SA 357  
 Db 386 AA 387  
 RESULT 8  
 PEP1\_CHICK STANDARD; PRT; 367 AA.  
 ID PEP1\_CHICK  
 AC P00793:  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Pepsin A precursor (EC 3.4.23.1).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]

```

RP SEQUENCE.
RX MEDLINE=84004412; PubMed=6617663;
RA Baudys M., Kostka V.;
RT "Covalent structure of chicken pepsinogen.";
RL Eur. J. Biochem. 136:89-99(1983).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR PIR: A00984; PCCH.
DR HSSP: P00794; 4CMS.
DR MEROPS: A01.001; -.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KM Hydrolyase: Aspartyl protease; Digestion; Stomach; Zymogen;
KW Glycoprotein; Gastric juice.
FT PROPEP 1 42
FT CHAIN 43 367
FT ACT_SITE 77 77
FT ACT_SITE 260 260
FT CARBOHYD 113 113
FT DISULFID 90 95
FT DISULFID 251 255
FT DISULFID 290 323
SQ SEQUENCE 367 AA; 40431 MW; 0C547E7ED8F5B341 CRC64;

Query Match 15.8%; Score 302; DB 1; Length 367;
Best Local Similarity 24.0%; Pred. No. 4.7e-18;
Matches 88; Conservative 69; Mismatches 125; Indels 84; Gaps 13;

QY 13 YVEMTVGSPPTQTLNIVDTGSSNFVGAAPHPFL-----HRYQRQLSSTYRDLKRG 65
DB 59 YGCTISIGTPQDFSVIFDTGSSNLVW---PSYICKSSACSNHKKRPDPKSSSTYVSTNET 115
QY 66 VVYPYTGKMEGELGTDLVSIHPGPNVTVRANIAITESDK-FFINGSMWEGITGLAYAE 124
DB 116 VVTAATGTSMSGLIGDYTAIV--SSIDVQNOQFGLSTEPGSEFFYCNFDGILGAFPS 172
QY 125 IARPDDSLPEFPDSLVKQTHV-PNLSLQLCGAGFPPLNSEVLASVSGSMITGIDHSLY 183
DB 173 IS--SSGATPPVDNMMSQHLVADLFVYLSKDG-----ETGSEVLEGGIDPNYNT 220
QY 184 TQSLWTPPIRREMYEVITVRVEINGODLK--MDCKEYNYDKSIYDSGTNLRKKVFE 241
DB 221 TKGIVWPLSAETWQIYMDRVTGKMYACFFTC-----QAIYDTGTSILVMQGAIVN 274
QY 242 AAVKSKAKASSNE-----KPPDGFMLGEQLVCMQAGTTPWNIFPVISLYLMGEVTQS 294
DB 275 RIHKDLGVSSDEISODDISKLDP-----VTFHNGHA---- 307
QY 295 FRITTLPOQYLRPEVDVATSDCYKFAISOSSTGT-----VMGAVIMEGGYVFPDAR 348
DB 308 -----FTLPASAVYINENSGCMGLGFENMGTFELGEOMLIGDVFIREYVIFDRAN 358
QY 349 KRIGFA 354
DB 359 NKVGLS 364

```

```

OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCB1_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
RC TISSUE=gastric mucosa;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL Eur. J. Biochem. 202:205-215(1991).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPsinogen GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- MISCELLANEOUS: EACH PEPsinogen IS CONVERTED TO CORRESPONDING
CC PEPsin AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA
CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59753; CA442425.1; -.
DR PIR: S16065; S16065.
DR PIR: S19682; S19682.
DR HSSP: P00790; IPSN.
DR MEROPS: A01.001; -.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001791; Laminin_1.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Multigene family; Signal; Glycoprotein.
FT SIGNAL 1 15
FT PROPEP 16 38
FT PROPEP 39 62
FT CHAIN 63 388
FT ACT_SITE 94 94
FT ACT_SITE 277 277
FT DISULFID 107 112
FT DISULFID 268 272
FT DISULFID 311 344
FT CARBOHYD 88 88
SQ SEQUENCE 388 AA; 41955 MW; A2923AB1F7CDEB9 CRC64;

Query Match 15.8%; Score 302; DB 1; Length 388;
Best Local Similarity 28.6%; Pred. No. 5.1e-18;
Matches 103; Conservative 58; Mismatches 131; Indels 68; Gaps 15;

QY 13 YVEMTVGSPPTQTLNIVDTGSSNFVGAAPHPFL-----HRYQRQLSSTYRDLKRG 65
DB 76 YFGTIGTGTPAQNTVTFDTGSSNLVW---PSYICKSSACSNHKKRPDPKSSSTYVSTNET 132
QY 66 VVYPYTGKMEGELGTDLVSIHPGPNVTVRANIAITESDK-FFINGSMWEGITGLAYAE 124
DB 133 VSIYTGTSMTGLIGDYTAIV--GGISDINQIFGLSTEPGSEFFYCNFDGILGAFPS 189
QY 125 IARPDDSLPEFPDSLVKQTHV-PNLSLQLCGAGFPPLNSEVLASVSGSMITGIDHSLY 183

```



Db 190 IS--SSGATPVDNINMQRIVSQDLFSVYLSD-----DQS-----GSVIFGQIDSSY 237

QY 184 TGSMTWPIREMYEYIYVAVELNGODL--KMDCKEYNDKSIYDSCGTNLRPKKVF 241

Db 238 TGSINWVPSVEGWQIVSDITNGKTIKACKC-----QALVDGTSLTNGPTSP 291

QY 242 AAVSIRKASSTEFKPDGFWLGEOLV-CWQAGTTPWNIFFPVISLYLMGEVYTNOSFRITIL 300

Db 292 NIGSDIGASEMD-----GEMVYSCSAISLPLDIYF-----TIN 325

QY 301 POQYLRPEVDVATISQDDCYK-----FAISSSTGCTWAGVIMGEFVYVDPARRKRIGFA 354

Db 326 GQYVPLPSPAYILIOSGCTSGFQGMVPTESGELWILGDVFIKQFTVFRANNGVLA 385

RESULT 10

PEPE\_CHICK STANDARD: PRT: 383 AA.

AC P16476:

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Embryonic pepsinogen precursor (EC 3.4.23.-).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

NCBI\_TaxID=9031;

NCBI [1]

SEQUENCE FROM N.A.

RP MEDLINE=86227903; PubMed-3131317;

RA Hayashi K., Agata K., Mochii M., Yasugi S., Eguchi G., Mizuno T.;

RT "Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsinogen: phylogenetic relationship with prothymosin.";

RT J. Biochem. 103:290-296(1988).

CC -1- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC

CC -1- PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: D00215; BAA00153.1; -

DR PIR: A11443.

DR HSSP: P00794; 4CMS.

DR MEROPS: A01.028;

DR InterPro: IPR001969; Asp\_protease.

DR InterPro: IPR001461; Pepsin.

DR Pfam: PF00026; asp. 1.

DR PRINTS: PR00792; PEPsin.

DR PROSITE: PS00141; Asp\_PROTEASE; 2.

DR HydroLase: Asparyl protease; Digestion; Signal; Glycoprotein.

KW SIGNAL

FT CHAIN 1 16

FT ACT\_SITE 17 383

FT ACT\_SITE 94 94

FT ACT\_SITE 276 276

FT DISULFID 107 112

FT DISULFID 267 271

FT DISULFID 310 344

FT CARBOHYD 132 132

FT CARBOHYD 204 204

FT CARBOHYD 309 309

FT CARBOHYD 350 350

FT CARBOHYD 51 51

FT VARIANT 383 AA; 41719 MW; 1642796871611F54 CRC64;

SEQUENCE

Query Match 15.7%; Score 298.5; DB 1; Length 383;

Best Local Similarity 25.2%; Pred. No. 9,8e-18;

Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14.

QY 13 YVEMTVGSPPOFLNLTVDGSSNFAVGA---APHPFLHRYQRLSTYDLRKGVY 68

Db 76 YGTSISIGPPQDFVTFVFDGSSNLWVPVSVCTSPKQSHQFNPSQSTYSTQGNLSI 135

QY 69 PYQGWKEGELGDLVSIHPGPVTVRANAIATTESDKFEINGSNMEGLGLAYAEIARP 128

Db 136 HGTGDMEGIVGCDYTVASIMDTNQLFGLST-SEPGQFFVY-VKFDGLIGLAPSLAA- 192

QY 129 DLSLEPPFDSLVRQTHV-PNLFSLQCGAGFPLNOSSEVLASVGSIMIGIDHSLTGS 187

Db 193 -DGTTPVDNMVNVESLLEQMLFSVYLS-----RPMGSMNVFGLIDESTYTGSI 240

QY 188 WYTPIREMYEYIYVAVELNGODL--KMDCKEYNDKSIYDSCGTNLRPKKVF 245

Db 241 NMIPVSYQGTWQISMDSIYVKNQEIACSSG-----QALIDGTSLVAPASDINDIOS 294

QY 246 SIKASSTEFKPDGFWLGEOLVCMQAGTTPWNIFFPVISL-----YLMGEVYTNOSFRITILP 301

Db 295 AVGANQNT-----YGEYSV-----NCSHILAMPDVVFIIGI----- 326

QY 302 QQYLRPEVDVA---TSQDDCYKRAISQSSNGTWGAVIMGEFVYVDPARRKRIGFA 354

Db 327 -QY--PVPALAYTBQNGQTCMSSSFQNSADWLILGDVFIKQFTVFRANNGVLA 380

RESULT 11

CATE\_HUMAN STANDARD: PRT: 396 AA.

ID CATE\_HUMAN

AC P14091;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Cathepsin E precursor (EC 3.4.23.34).

DE CTSE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OC NCBI\_TaxID=9606;

NCBI [1]

SEQUENCE FROM N.A.

RP MEDLINE=89380302; PubMed-2674141;

RA Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;

RT "Human gastric cathepsin E. Predicted sequence, localization to

RT chromosome 1, and sequence homology with other aspartic

RT proteinases.";

RT J. Biol. Chem. 264:16748-16753(1989).

RL

RP SEQUENCE FROM N.A.

RP MEDLINE=92112877; PubMed-1370478;

RA Azuma T., Liu W.G., Vander laan D.J., Bowcock A.M., Taggart R.T.;

RT "Human gastric cathepsin E gene. Multiple transcripts result from

RT alternative polyadenylation of the primary transcripts of a single

RT gene locus at 1q31-q32.";

RL J. Biol. Chem. 267:1609-1614(1992).

RL

RP SEQUENCE FROM N.A.

RA Tatnell P.J., Kay J.;

RT "Human procathepsin E.";

RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.

RL

RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.

RP MEDLINE=90241267; PubMed-2334440;

RA Athauda S.B.P., Matsuzaki O., Kgeyama T., Takahashi K.;

RT "Structural evidence for two isozymic forms and the carbohydrate

RT attachment site of human gastric cathepsin E.";

RL Biochem. Biophys. Res. Commun. 168:878-885(1990).

CC -1- FUNCTION: DOE OF ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN

CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.

CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader

CC specificity.



```
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84424; AAA52300.1; -
DR EMBL; M84413; AAA52300.1; JOINED.
DR EMBL; M84417; AAA52300.1; JOINED.
DR EMBL; M84418; AAA52300.1; JOINED.
DR EMBL; M84419; AAA52300.1; JOINED.
DR EMBL; M84420; AAA52300.1; JOINED.
DR EMBL; M84421; AAA52300.1; JOINED.
DR EMBL; M84422; AAA52300.1; JOINED.
DR EMBL; J05036; AAA52300.1; JOINED.
DR EMBL; AJ250717; CAB82850.1; -
DR PIR; A34401; A34401.
DR PIR; A34643; A34643.
DR PIR; A42038; A42038.
DR HSSP; P00784; 4CMS.
DR MEROPS; A01.010; -.
DR MIM; I16890; -.
DR InterPro; IPR001969; Asp-protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; Pepsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydroxylase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 17
FT PROPEP 18 53
FT CHAIN 54 396
FT MOD_RES 18 18
FT ACT_SITE 96 96
FT ACT_SITE 281 281
FT DISULFID 60 60
FT DISULFID 109 114
FT DISULFID 272 276
FT DISULFID 314 351
FT CARBOHYD 90 90
FT CARBOHYD 220 220
FT CARBOHYD 333 333
SQ SEQUENCE 396 AA; 42793 MW; 40B643C5FB01521E CRC64;

Query Match 15.7%; Score 298.5; DB 1; Length 396;
Best Local Similarity 26.7%; Pred. No. 1e-17;
Matches 96; Conservative 61; Mismatches 141; Indels 61; Gaps 14;

OY 13 YVEKTVGSPPTLNIIVDTGSSNFVAVG----APHPLRHRYROLSTYRDLRKGVYV 68
DB 78 YFGRTISIPSPONFVIPTDGTSSNLMWVSVCYCTSPACKTHSRFPQSQSTISQGGQSSSI 137
OY 69 PYTGKMEGELGTDIVSRPHRPNTVVRANITATESDKFFINGSNWEGICGLAVETARP 128
DB 138 OYGTGSLSGITIGADGVSV-EGVIYVGQFGSGSVTEPGTFVD-AEFOGILGLGYPSTLA- 193
OY 129 DQSLSEFFDSLVKQTHVNPFLSLQCGAGFPLNQSEVLAVSGVMIGIGIDHSYTGSLW 188
DB 194 VGGVYVPEVDNNMAQ---NLVDLPMEFSVYMSNPE---GGAGSELITGGIDHSHFSSSLN 246
OY 189 YTPIRREMYEVIIVRVINQDLKMDCKEYNYKSYVDSGTNLRLPKRYFEAAVYSIK 248
DB 247 WVPYTKAWIWDIADNLOVGG--TVMFCS--GCQAIYDGTSLITGSPDKIKOLQNAIG 302
OY 249 AASSTKPPDFMLGELQVLCWQAGTTPMNIPIVSLVLMGEVTVNQSRITIIIPQOYLAPV 308
DB 303 AAP-----VDSGEVAVE-----CANLWNPADVTFTING-----VPYLSPTAY--TL 341
```

```
OY 309 EDVATNSQDDCYKFAISQSTG-----TWGAVIMEGFYVVEDARRRIGFA 354
DB 342 LDFVDGMQFC-----SSGFGQGDHPHPAGPLMIITLDVFIROPYSVFDGNNRVGIA 392

RESULT 12
PEP2_RABIT
ID PEP2_RABIT STANDARD: PRT: 387 AA.
AC P27821.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin II-2/3 precursor (EC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NM NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59235; AAA83369.1; -
DR PIR; C38302; C38302.
DR HSSP; P00790; 1PSN.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Asp-protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; Pepsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydroxylase; Aspartyl protease; Digestion; Stomach; Gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59
FT CHAIN 60 387
FT MOD_RES 129 129
FT ACT_SITE 93 93
FT ACT_SITE 276 276
FT DISULFID 106 111
FT DISULFID 267 271
FT DISULFID 310 343
SQ SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;

Query Match 15.5%; Score 296; DB 1; Length 387;
Best Local Similarity 26.9%; Pred. No. 1.6e-17;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

OY 13 YVEKTVGSPPTLNIIVDTGSSNFVAVGAPHPF-----LHRYQROLSTYRDLKRG 65
DB 75 YFGRTISIGTPPDVFIPTDGTSSNLMW---PSTGCSLACALHKKRNPEDSSSTYGTSET 131
```

QY 66 YVVPYTGKWEGLCTDVLSPHGPNTVRANIAITESDKFETNGSMEGILGLAYAEI 125  
 Db 132 LSTITGTSMTGILCYDVKVKSIEDTQIIGLKTESLFLF--APFDGILGLAPSI 189  
 QY 126 ARPDDSLPEFDSLVKQYH-PNLFSLQLCAGPPLNOSVLAISVGSMTIGIDHSIXT 184  
 Db 190 SSSDAT--PVEDNMWNEGLVSQDLFSVYLSSD-----EKGSLVMFGGIDSSYTT 237  
 QY 185 GSWPTPIREKRYEVIVREINGODLKM--DCKEYNDKSIYDSGTNTLRLEPKVPEA 242  
 Db 238 GSLNWPVSYEYEWQIYDVSINGETIACADSC-----QAIYDTGSLTGP-----TS 287  
 QY 243 AVKSIKASSTKFPDGLGEOLV-CWQAGTTPWNIPFVLSLYMGVYNOSEFRTILP 301  
 Db 288 AISNIOSTYGAK-----NLGENVISCASIDSLPDIIV-----TING 325  
 QY 302 QOYLPRVEDVATSDCKKFAISQSTG-----YKNAVIMEGYVEDRARKRGIFAV 355  
 Db 326 IYPIIPASAYILKEDDCTSGLEGNNVDITYGELMIGDVFIRGYFVFDNRANOLGLAA 385  
 QY 356 S 356  
 Db 386 A 386

## RESULT 13

CATE\_CAVPO STANDARD: PRT: 391 AA.

AC P25796;  
 DT 01-MAY-1992 (rel. 22, Created)  
 DT 01-MAY-1992 (rel. 22, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)  
 DE Cathepsin E precursor (EC 3.4.23.34).  
 GN CTSE.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.  
 NCBI\_Taxid=10141;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=92355614; PubMed=1644829;  
 RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koiwai O.,  
 Tanji M., Yakabe E., Athuda S.B., Takahashi K.;  
 RT "Gastric procathepsin E and progastricsin from guinea pig.  
 RT Purification, molecular cloning of cDNAs, and characterization of  
 RT enzymatic properties, with special reference to procathepsin E.";  
 RL J. Biol. Chem. 267:16450-16459(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=gastric mucosa;  
 RX MEDLINE=96073637; PubMed=8540321;  
 RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,  
 Tanji M., Athuda S.B., Takahashi K.;  
 RT "Isolation, characterization, and structure of procathepsin E and  
 RT cathepsin E from the gastric mucosa of guinea pig.";  
 RL Adv. Exp. Med. Biol. 362:211-221(1995).  
 CC -1- FUNCTION: DUE OF ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN  
 CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.  
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader  
 CC specificity.  
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M88653; AAA37052.1; -  
 DR EMBL: S80547; AAB35844.1; -  
 DR PIR: A43356; A43356.  
 DR HSSP: P00794; 4CMS.  
 DR MEROPS: A01.010; -  
 DR Interpro: IPR001969; Asp-protease.  
 DR Interpro: IPR001461; Pepsin.  
 DR Pfam: PR00026; asp.1.  
 DR PRINTS: PRO0792; PEPSTN.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolyase; Aspartyl protease; Glycoprotein; zymogen; signal.  
 FT SIGNAL.  
 FT PROPEP 1 18  
 FT CHAIN 54 391  
 FT ACT\_SITE 92 92  
 FT ACT\_SITE 276 276  
 FT DISULFID 56 56  
 FT DISULFID 105 110  
 FT DISULFID 267 271  
 FT DISULFID 309 346  
 FT CARBOHYD 86 86  
 SQ SEQUENCE 391 AA; 42132 MW; 78D216BF8C6DABD CRC64;  
 Query Match 15.5%; Score 296; DB 1; Length 391;  
 Best Local Similarity 26.9%; Pred. NO. 1.6e-17;  
 Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

QY 13 YVVPYTGKWEGLCTDVLSPHGPNTVRANIAITESDKFETNGSMEGILGLAYAEI 68  
 Db 74 YFGTISGSPQNTFVITDVTSSNLMWPSVCTSPACQTHVFPRLSSTRVRENGSFSI 133  
 QY 69 PYTGKWEGLCTDVLSPHGPNTVRANIAITESDKFETNGSMEGILGLAYAEI 128  
 Db 134 QYGTSLGILIGADQVY-EGILTVGQGFGEVQEPKTFVH-AEFDGILGLAPSLA- 190  
 QY 129 DLSLPEFDSLVKQYHVNLSLQLCAGPPLNOSVLAISVGSMTIGIDHSIXT 188  
 Db 191 -GGVTPVDNMMAQ-----NLVALPM-----FSYMSNPNCGSGSELTRGCTPSHFSSLN 241  
 QY 189 YPIREKRYEVIVREINGODLKMCKEYNDKSIYDSGTNTLRLEPKVPEA 248  
 Db 242 WVPVTKQAWQALDGIQV--DSVMFCSE--GCQAIYDTGSLTGP-----PKIKOLQ 293  
 QY 249 AASTKFPDGLGEOLV-CWQAGTTPWNIPFVLSLYMGVYNOSEFRTILP 303  
 Db 294 EALGATVYDEG-----SYOC-----ANLMMLDVT-----FLINGVPTLNTFA 333  
 QY 304 YLRPVEDVATSDCKKFAISQSTG-----TWGAVIMEGYVEDRARKR 350  
 Db 334 Y--TLIDFVDGMQV-----STGFEGLEIQTTPAGPLMIGDVFIRGYFVFDNR 383  
 QY 351 IGFA 354  
 Db 384 VGIA 387

## RESULT 14

CATD\_HUMAN STANDARD: PRT: 412 AA.

AC P07339;  
 DT 01-APR-1988 (rel. 07, Created)  
 DT 01-APR-1988 (rel. 07, Last sequence update)  
 DT 01-MAR-2002 (rel. 41, Last annotation update)  
 DE Cathepsin D precursor (EC 3.4.23.5).  
 GN CTSD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85270436; PubMed=3927292;  
 RA Faust P.L., Kornfeld S., Chirgwin J.M.;

RT "Cloning and sequence analysis of cDNA for human cathepsin D.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87231068; PubMed-3588310;  
 RA Westley B.R., May F.E.B.;  
 RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive  
 human breast cancer cells.";  
 RL Nucleic Acids Res. 15:3773-3786(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91299158; PubMed-2069717;  
 RA Redeker B., Heckendorf B., Grosch H.W., Wersmann G., Hasilik A.;  
 RT "Molecular organization of the human cathepsin D gene.";  
 RL DNA Cell Biol. 10:423-431(1991).  
 RN [4]  
 RP SEQUENCE OF 1-22 FROM N.A.  
 RX MEDLINE-94085791; PubMed-8262386;  
 RA May F.E., Smith D.J., Westley B.R.;  
 RT "The human cathepsin D-encoding gene is transcribed from an estrogen-  
 regulated and a constitutive start point.";  
 RL Gene 134:277-282(1993).  
 RN [5]  
 RP SEQUENCE OF 1-22 FROM N.A.  
 RX MEDLINE-95021301; PubMed-7935485;  
 RA Augereau P., Miralles F., Cavailles V., Gaudeliet C., Parker M.,  
 RT Rochefort H.;  
 RL "Characterization of the proximal estrogen-responsive element of  
 human cathepsin D gene.";  
 RL Mol. Endocrinol. 8:693-703(1994).  
 RN [6]  
 RP SEQUENCE OF 170-180.  
 RC TISSUE=Liver;  
 RA Hochstrasser D.F., Fritiger S., Paquet N., Bairoch A., Ravier F.,  
 RA Pasquell C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargias R.,  
 RA Appel R.D., Hughes G.J.;  
 RL Submitted (JUN-1992) to the SWISS-PROT data bank.  
 RN [7]  
 RP VARIANT VAL-58.  
 RX MEDLINE-20179010; PubMed-10716266;  
 RA Pappasotiropoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,  
 RA Maier W., Pauls J., Lautenschlager N., Heun R.;  
 RT "A genetic variation of cathepsin D is a major risk factor for  
 Alzheimer's disease.";  
 RL Ann. Neurol. 47:399-403(2000).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).  
 RC TISSUE=Spleen;  
 RX MEDLINE-9323670; PubMed-8467789;  
 RA Metcalf P., Fusek M.;  
 RT "Two crystal structures for cathepsin D: the lysosomal targeting  
 signal and active site.";  
 RL EMBO J. 12:1293-1302(1993).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RC TISSUE=Liver;  
 RX MEDLINE-93342076; PubMed-8393577;  
 RA Baldwin E.T., Bhat T.N., Gulnik S., Houser M.V., Sowder R.C. II,  
 RA Cachau R.E., Collins J., Silva A.M., Erickson J.W.;  
 RT "Crystal structures of native and inhibited forms of human cathepsin  
 D: Implications for lysosomal targeting and drug design.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).  
 CC -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR  
 CC PROTEIN BREAKDOWN.  
 CC -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,  
 CC that of pepsin A. Does not cleave the 4-Gln-I-His-5 bond in B  
 CC chain of insulin.  
 CC -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- POLYMORPHISM: The Val-58 allele is significantly overrepresented  
 CC in demented patients (11.8%) compared with nondemented controls  
 CC (4.9%). Carriers of the Val-58 allele have a 3.1-fold increased  
 CC risk for developing AD than noncarriers.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EURARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M11233; AAB59529.1; -;  
 CC EMBL; X05344; CA828855.1; -;  
 CC EMBL; M63138; AAB51922.1; -;  
 CC EMBL; M63134; AAB51922.1; JOINED.  
 CC EMBL; M63135; AAB51922.1; JOINED.  
 CC EMBL; M63136; AAB51922.1; JOINED.  
 CC EMBL; M63137; AAB51922.1; JOINED.  
 CC EMBL; L12980; AAB16314.1; -;  
 CC EMBL; S74689; AAD14156.1; -;  
 CC EMBL; S52557; AAD13868.1; -;  
 CC PIR; A25771; KHHUD.  
 CC PDB; 1LYA; 31-JAN-94.  
 CC PDB; 1LYB; 31-JAN-94.  
 CC MEROPS; A01.009; -;  
 CC SWISS-2DPAGE; P07339; HUMAN.  
 CC Sienra-2DPAGE; P07339; -;  
 CC MIM; 116840; -;  
 CC InterPro; IPR001969; Asp-protease.  
 CC InterPro; IPR001461; Pepsin.  
 CC Pfam; PF00026; asp. 1.  
 CC PRINTS; PRO0792; PEPsin.  
 CC PROSITE; PS00141; ASP\_PROTEASE; 2.  
 CC HydroLase; Aspartyl protease; Glycoprotein; Lysosome; Signal; zymogen;  
 CC K0 Polymorphism; Alzheimer's disease; 3D-structure.  
 CC KW SIGNAL 1 18  
 CC FT PROPEP 19 64 ACTIVATION PEPTIDE.  
 CC FT CHAIN 65 412 CATHEPSIN D.  
 CC FT CHAIN 65 161 LIGHT CHAIN (PROBABLE).  
 CC FT ACT\_SITE 169 412 HEAVY CHAIN (PROBABLE).  
 CC FT ACT\_SITE 97 97  
 CC FT DISULEPID 110 117  
 CC FT DISULEPID 110 117  
 CC FT DISULEPID 286 290  
 CC FT DISULEPID 329 366  
 CC FT CARBOHYD 134 134  
 CC FT CARBOHYD 263 263  
 CC FT VARIANT 58 58  
 CC FT  
 CC FT STRAND 67 74  
 CC FT TURN 75 77  
 CC FT STRAND 78 85  
 CC FT TURN 86 89  
 CC FT STRAND 90 97  
 CC FT TURN 98 99  
 CC FT STRAND 103 107  
 CC FT TURN 108 109  
 CC FT TURN 112 113  
 CC FT HELIX 115 118  
 CC FT TURN 119 119  
 CC FT STRAND 123 123  
 CC FT HELIX 125 127  
 CC FT TURN 129 130  
 CC FT STRAND 132 141  
 CC FT TURN 146 158  
 CC FT STRAND 172 184  
 CC FT HELIX 188 192  
 CC FT STRAND 197 200  
 CC FT HELIX 204 206  
 CC  
 CC N-LINKED (GLCNAC. . .).  
 CC A -> V (ASSOCIATED WITH INCREASED RISK IN  
 CC AD; POSSIBLY INFLUENCES SECRETION AND  
 CC INTRACELLULAR MATURATION; IN  
 CC DBSNP:175711.  
 CC /FTID=VAR\_011621.

FT	HELIX	208	210
FT	HELIX	214	220
FT	TURN	221	222
FT	STRAND	228	233
FT	STRAND	243	247
FT	TURN	248	248
FT	HELIX	252	254
FT	STRAND	255	263
FT	STRAND	267	267
FT	TURN	268	269
FT	STRAND	270	279
FT	TURN	280	281
FT	STRAND	284	285
FT	TURN	287	288
FT	STRAND	290	294
FT	TURN	296	297
FT	STRAND	298	298
FT	STRAND	301	303
FT	HELIX	305	315
FT	TURN	316	316
FT	STRAND	318	319
FT	TURN	322	323
FT	STRAND	325	328
FT	HELIX	329	334
FT	STRAND	338	342
FT	TURN	343	344
FT	STRAND	345	349
FT	HELIX	351	354
FT	STRAND	355	357
FT	TURN	359	362
FT	STRAND	365	368
FT	STRAND	370	372
FT	TURN	377	379
FT	STRAND	383	385
FT	HELIX	387	390
FT	TURN	391	392
FT	STRAND	393	398
FT	TURN	399	402
FT	STRAND	403	409

[illegible]

RESULT 15	PEP4_RABIT	STANDARD:	PRT:	387 AA.
AC	P28713;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, last annotation update)			
DE	Pepsin II-4 precursor (EC 3.4.23.1) (Pepsin A).			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID:9986;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91009127; PubMed=2129536;			
RA	Kageyama T., Tanabe K., Koike O.;			
RT	"Structure and development of rabbit pepsinogens. Stage-specific			
RT	zymogens, nucleotide sequences of cDNAs, molecular evolution, and			
RT	gene expression during development."			
RL	J. Biol. Chem. 265:17031-17038(1990).			
CC	-I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS			
CC	INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE			
CC	ALSO CLEAVED TO SOME EXTENT.			
CC	-I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-,			
CC	-I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP II, III AND III WHERE			
CC	THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.			
CC	-I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY			
CC	HORMONES AND RELATED SUBSTANCES.			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE			
CC	EUKARYOTIC ASPARTYL PROTEASES FAMILY.			
CC	PIR: D38302; D38302.			
DR	HSSP: P00790; 1PSN.			
DR	MEROPS: A01.001; -.			
DR	InterPro: IPR001969; Asp-protease.			
DR	InterPro: IPR001461; Pepsin.			
DR	Pfam: PF00026; asp_1.			
DR	PRINTS: PR00792; PEPIN.			
DR	PROSITE: PS00141; Asp-PROTEASE. 2.			
KW	Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;			
KW	Zymogen; Signal; Phosphorylation; Multigene family.			
FT	SIGNAL	1	15	
FT	PROPEP	16	59	ACTIVATION PEPTIDE.
FT	CHAIN	60	387	PEPSIN II-4.
FT	MOD_RES	129	129	PHOSPHORYLATION (POTENTIAL).
FT	ACT_SITE	93	93	BY SIMILARITY.
FT	ACT_SITE	276	276	BY SIMILARITY.
FT	DISULFID	106	111	BY SIMILARITY.
FT	DISULFID	267	271	BY SIMILARITY.
FT	DISULFID	310	343	BY SIMILARITY.
FT	SEQUENCE	387 AA;	42052 MW;	21ADDD07782A89585 CRC64;
QY	Query Match	15.5%;	Score 295;	DB 1; Length 387;
	Best Local Similarity	26.1%;	Pred. No. 2e-17;	
	Matches 97;	Conservative 65;	Mismatches 123;	Indels 86; Gaps 14.
Db	13 YYAMTWGSPQTLNITLVDSGSSNEAVGAPHP-----LHRYTORQLSTYRDLRKG 65	1	11	1
QY	1	1	1	1
Db	75 YFGTISIGTPQDDTYFTFGSSLMWY---PSTYCSSLALAKRRNPEDSSYQGTST 131	1	1	1
QY	1	1	1	1
Db	66 VYVYTGKKEGELGTDLVSIIPHGPNTVRANAIATESDKFF-----INSGNWE 115	1	1	1
QY	1	1	1	1
Db	132 LSTIYGGSGMGLIGYTV-----KGSIEDTNGIFGLSKTEPGLFLPAPFD 179	1	1	1
QY	1	1	1	1
Db	116 GILGLAVAEIARPDLSLEPPFDSLVKQTHV-PNLFSLQLGAGPPLNQSIVLASVGSMI 174	1	1	1
QY	1	1	1	1
Db	180 GILGLAVPSSISSDAT--PVFDNNMNGVLVSQDLSVYLLSSD-----EKSGIVM 227	1	1	1
QY	1	1	1	1
Db	175 IGGIDHSLYTGSSLMWTPPIRREMYEVLIVREINIGODLKM--DCKEYNDKSIYDSGTTN 232	1	1	1
QY	1	1	1	1
Db	228 FGGIDSSITYSGLMKNVPSYGYQIIMDSVINGETIACADSC-----QAIYDTGTSI 281	1	1	1
QY	1	1	1	1
Db	233 LIRKKVEAVKSIKASITEKPEPDGFWLGEOLV-CWQAGTTPWNIPIVSIYLMEVTV 291	1	1	1

Db 282 LTGP---TSAISNIOSTIGASK---NLGENVISCSAIDSLPDIVF----- 321  
QY 292 NOSFRITILPOQYLRPVEDVATSQDDCYKFAISOSTGT-----VMGAVIMEGFYVFD 345  
Db 322 -----TINGIQYPLPASAVILKEDDCTSGLEGMMVDYTGELMILGDVFIQYFTVFD 375  
QY 346 RARRRIGFANVS 356  
Db 376 RANNQGLLAAA 386

Search completed: October 30, 2002, 12:27:48  
Job time : 32.0442 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 112.646 seconds  
(without alignments)  
554,401 Million cell updates/sec

Title: US-09-724-571-75

Perfect score: 1907  
Sequence: 1 MVDNLRKSGGGYIVEMTVG.....VVEDRRKRKIGFAVSACHVH 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1907	100.0	532	4 Q9ULS1	Q9ULS1 homo sapien
2	1752.5	91.9	476	4 Q9BYC1	Q9BYC1 homo sapien
3	1649	86.5	457	4 Q9BYC0	Q9BYC0 homo sapien
4	1504.5	78.9	432	4 Q9BYB9	Q9BYB9 homo sapien
5	1106	58.0	439	4 Q9H2V8	Q9H2V8 homo sapien
6	1088	57.1	514	11 Q9JIL8	Q9JIL8 mus musculi
7	996	52.2	266	11 Q9CUD5	Q9CUD5 mus musculi
8	956.5	50.2	396	4 Q9NZL1	Q9NZL1 homo sapien
9	907	47.6	468	4 Q9PDL2	Q9PDL2 homo sapien
10	712.5	37.4	213	4 Q9PDL2	Q9PDL2 homo sapien
11	547.5	28.7	255	11 Q9KIP7	Q9KIP7 mus musculi
12	332.5	17.4	391	5 Q9VKP6	Q9VKP6 drosophila
13	332	17.4	354	5 Q9GIX7	Q9GIX7 boophilus m
14	308	16.2	384	13 Q9DEC4	Q9DEC4 xenopus lae
15	308	16.2	385	13 Q9DEC4	Q9DEC4 rana catesb
16	305	16.0	386	6 Q9GMV7	Q9GMV7 rhinolophus

17	305	16.0	387	6 Q9GMV8	Q9GMV8 scorex ungui
18	304.5	16.0	372	5 Q9VLK3	Q9VLK3 drosophila
19	304.5	16.0	386	6 Q9BGU5	Q9BGU5 bos taurus
20	304	15.9	387	6 Q9GMV9	Q9GMV9 suncus muri
21	304	15.9	387	13 Q9DDV5	Q9DDV5 salvelinus
22	303	15.9	383	13 Q9DEC3	Q9DEC3 xenopus lae
23	303	15.9	383	13 Q9DE45	Q9DE45 salvelinus
24	301	15.8	382	13 Q9PRG9	Q9PRG9 gallus gall
25	299	15.7	423	5 Q9VKP7	Q9VKP7 drosophila
26	298	15.6	376	13 Q9PDR8	Q9PDR8 pseudopleur
27	295.5	15.5	384	13 Q9I322	Q9I322 rana catesb
28	294	15.4	396	6 Q9J428	Q9J428 chionodraco
29	290.5	15.2	381	6 Q9GK11	Q9GK11 camelus dro
30	289	15.0	386	6 Q9GMV6	Q9GMV6 canis famli
31	287	15.0	399	13 Q9J458	Q9J458 canis famli
32	285.5	15.0	444	5 Q21966	Q21966 caenorhabdi
33	284	14.9	398	13 P87370	P87370 oncorhynch
34	280.5	14.7	427	5 P91802	P91802 schistosoma
35	279.5	14.7	390	6 Q9GK10	Q9GK10 camelus dro
36	278	14.6	370	6 Q9TWL1	Q9TWL1 bos taurus
37	278	14.6	378	13 Q9PDR9	Q9PDR9 pseudopleur
38	278	14.6	399	13 Q9PDR9	Q9PDR9 pseudopleur
39	278	14.6	422	5 Q96906	Q96906 brachydanio
40	277.5	14.6	380	6 Q28950	Q28950 sus scrofa
41	276	14.5	396	13 Q9DEX3	Q9DEX3 clupea hare
42	275	14.4	385	6 Q29080	Q29080 sus scrofa
43	275	14.4	413	3 Q14413	Q14413 pichia angu
44	275	14.4	446	5 Q9N9H3	Q9N9H3 necator ame
45	272	14.3	389	13 Q9PWK1	Q9PWK1 gallus gall

## ALIGNMENTS

### RESULT 1

ID Q9ULS1 PRELIMINARY; PRT; 532 AA.  
AC Q9ULS1;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE KIAA1149 PROTEIN (FRAGMENT).  
GN KIAA1149.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20039618; PubMed=10574461;  
RA Hirosewa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;  
RT "Characterization of cDNA clones selected by the Genemark analysis  
from size-fractionated cDNA libraries from human brain.";  
RL DNA Res. 6:329-336(1999).  
DR EMBL; AB032975; BAA86463.2; -  
DR HSSP; P56272; 1AM5.  
DR MEROPS; A01.004; -  
DR InterPro; IPR001969; Asp-protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF000026; asp; 1.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
FT NON TER 1  
SQ SEQUENCE 532 AA; 58720 MW; 98B135D0DSFBD2E8 CRC64;

Query Match 100.0%; Score 1907; DB 4; Length 532;

Best Local Similarity 100.0%; Pred. No. 6e-159;

Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVDNLRKSGGGYIVEMTVGSPQTLNIIIVDSSNFAYGAAPHLRHRYORQLSSYTR 60  
DB 94 MVDNLRKSGGGYIVEMTVGSPQTLNIIIVDSSNFAYGAAPHLRHRYORQLSSYTR 153

Qy	61	DLRKGVVPTTQGMBSGLDLDVSTPHGNVNVANIAITSSDKFPLNGSWBELLG	120
Db	154	DLRGGVVPTTQGMBSGLDLDVSTPHGNVNVANIAITSSDKFPLNGSWBELLG	213
Qy	121	AYAIARPDSDLPEFPDGLVKOTHPVPLFSLQCGAGFPLNDSGLVASVGSMLIGIDH	180
Db	214	AYAIARPDSDLPEFPDGLVKOTHPVPLFSLQCGAGFPLNDSGLVASVGSMLIGIDH	273
Qy	181	SLYGSIMTYPIRREWEYEVIIIVREINGDLMKDCKEYNIDKSIYDSGTTNLRPKYF	240
Db	274	SLYGSIMTYPIRREWEYEVIIIVREINGDLMKDCKEYNIDKSIYDSGTTNLRPKYF	333
Qy	241	EAAVKSITKAASSTKFPDGMFLGDLVCMQAGTTPMNIIFVYISLYLMGEVTTNOSFRITLL	300
Db	334	EAAVKSITKAASSTKFPDGMFLGDLVCMQAGTTPMNIIFVYISLYLMGEVTTNOSFRITLL	393
Qy	301	POQYLRPEVEDVATSDDDCYKFAISOSSTGVMGVIINEGFVYVDDRARRKIRGFVAVSACHV	360
Db	394	POQYLRPEVEDVATSDDDCYKFAISOSSTGVMGVIINEGFVYVDDRARRKIRGFVAVSACHV	453
Qy	361	H 361	
Db	454	H 454	

RESULT 2	
09BYC1	PRT; 476 AA.
PRELIMINARY;	
ID 09BYC1	
AC 09BYC1	
DT 01-JUN-2001	(REMBLrel. 17, Created)
DT 01-JUN-2001	(REMBLrel. 17, Last sequence update)
DT 01-DEC-2001	(REMBLrel. 19, Last annotation update)
DE	BETA-SITE APP CLEAVING ENZYME I-476.

OC Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RC MEDLINE=21408467; PubMed=11516562;  
RA Tanahashi H., Tabita T.;  
RT "Three novel alternatively spliced isoforms of the human beta-site APP  
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide  
RT production.";  
RL Neurosci. Lett. 307:9-12(2001).  
RL EMBL, AB050436, BAB40931.1; -.  
DR HSSP; P32329; 1YPS.  
DR InterPro; IPR001969; Asp.protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF000026; asp. 1.  
DR PRINTS; PR00792; PEPSTN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1  
SQ SEQUENCE 476 AA; 52907 MW; 6C8C67F8A953AF66 CRC64;

Query Match	91.9%	Score 1752.5	-DB 4	Length 476
Best Local Similarity	93.1%	Pred. No. 145		
Matches 336	Conservative	0	Mismatches 0	Indels 25
				Gaps 1

QY 1 MVDNLGRKSGGQYVEEMTVGSPQTLNLIVDTGSSNFVAAGAAPHFLHRYQROLSTYY 60

Db 63 MVDNLGRKSGGQYVEEMTVGSPQTLNLIVDTGSSNFVAAGAAPHFLHRYQROLSTYY 122

QY 61 DLRKGYVYVYTGKMEGLGTDVLSIPHGPNVYVRANIAATTEDEKFFINGSNMEGLGL 120

Db 123 DLRKGYVYVYTGKMEGLGTDVLSIPHGPNVYVRANIAATTEDEKFFINGSNMEGLGL 182

QY 121 AAALAIRPDDSLPEFFDLSVKQTHPNLNLQLTGAGGPLNQSEVLSVGSMTIGSIDH 180

Db 183 AAALAIR-----LCGAGPLNQSEVLSVGSMTIGSIDH 217

QY	181	SLYGTSLWYTPPIREEMYYEVIYIAVEINQDLDKMDCKEYVYKDSIYDNGTTLRLPKKVF	240
Db	218	SLYGTSLWYTPPIREEMYYEVIYIAVEINQDLDKMDCKEYVYKDSIYDNGTTLRLPKKVF	277
QY	241	EAAYKSIKAASSTKEKFPDGFWLGEOIYCWQAGTTPNNIIPVYISLYLMEGVYNSFRITLL	300
Db	278	EAAYKSIKAASSTKEKFPDGFWLGEOIYCWQAGTTPNNIIPVYISLYLMEGVYNSFRITLL	337
QY	301	POQYLRPEVEDYATSDODCYKFAISQSSTGYTMGAVIMEGFVYVDRKRRKIGFAVSACHV	360
Db	338	POQYLRPEVEDYATSDODCYKFAISQSSTGYTMGAVIMEGFVYVDRKRRKIGFAVSACHV	397
QY	361	H 361	
Db	398	H 398	

RESULT 3		
09BYC0		
ID	09BYC0	PRELIMINARY; PRT: 457 AA.
AC	09BYC0	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 15, Last annotation update)	
DE	BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME TYPE C).	
DE	GN	
GN	BACE.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
OX	NCBI_TaxID=9606;	

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=21408467; Pubmed=11516562;  
 RA Tanahashi H., Tablita T.;  
 RT "Three novel alternatively spliced isoforms of the human beta-site APP  
 RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide  
 RT production.";  
 RL Neurosci. Lett. 307:9-12(2001).  
 RN [2]

RP SEQUENCE FROM N.A.  
RA TISSUE=EXOCRINE PANCREAS;  
RC Zaccchetti D., De Pietri Tonelli D., Schnurbus R.;  
RT "New beta-site APP cleaving enzyme isoform (BACE-1c) obtained from  
RT human pancreas.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB050437; BAB40932.1; -  
DR EMBL: AF338817; AAK38375.1; -  
DR HSSP: P32329; 1YPS  
DR InterPro: IPR001969; Asp-protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; asp; 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN 1.  
DR PROSITE: 457 AA; 51068 MW; C794C9A9E855E7A2 CRC64;  
SQ SEQUENCE

Query Match	86.5%;	Score 1649;	DB 4;	Length 457;
Best Local Similarity	87.8%;	Pred. No. 2.2e-136;		
Matches 317;	Conservative	0;	Mismatches	0;
			Indels	44;
			Gaps	1.

QY 1 MVDNLGRKSGSGGYVAMTGGSPQRLNLIVPDGSSNENFVAGAAPPLFLHRYQRLSSTYR 60

Db 63 MVDNLGRKSGSGGYVAMTGGSPQRLNLIVPDGSSNENFVAGAAPPLFLHRYQRLSSTYR 122

QY 61 DLRRGYVVPRTQSGKMEGELGDTLVSIPHGPNVYRANIAATTESDKFFINSMEWEGILGL 120

Db 123 DLRRGYVVPRTQSGKMEGELGDTLVSIPHGPNVYRANIAATTESDKFFINSMEWEGILGL 145

QY 121 AVAELIARPDLSLEPFEDSVYKQTHVNNFLSLDLCAGEPINQSEVTLASVGGSMILGGIDH 180

Db 146 -----PDSLEPFEDSLVKQTHVNNFLSLDLCAGEPINQSEVTLASVGGSMILGGIDH 198



QY 181 SLVGTSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDKSIYDSGTTNLRPKYF 240  
 Db 199 SLVGTSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDKSIYDSGTTNLRPKYF 258  
 QY 241 EAAVKSIAASTETKPFDPGFWLGEOLVCMQAGTTPMNIFFVLSLYLMGEVYTNOSFRITIL 300  
 Db 259 EAAVKSIAASTETKPFDPGFWLGEOLVCMQAGTTPMNIFFVLSLYLMGEVYTNOSFRITIL 318  
 QY 301 POQYLRPEVEDVATSDODCKYKFAISQSTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 360  
 Db 319 POQYLRPEVEDVATSDODCKYKFAISQSTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 378  
 QY 361 H 361  
 Db 379 H 379

## RESULT 4

Q9BYB9 PRELIMINARY; PRT; 432 AA.  
 AC Q9BYB9;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE BETA-SITE APP CLEAVING ENZYME I-432.  
 GN BACE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRIN;  
 RX MEDLINE=21408467; PubMed=11516562;  
 RA Tanahashi H., Tabira T.;  
 RT "Three novel alternatively spliced isoforms of the human beta-site APP  
 RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide  
 RT production.";  
 RL Neurosci. Lett. 307:9-12(2001).  
 DR EMBL; AB050438; BAB40933.1; -.  
 DR HSSP; P32329; IYPS.  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 SQ SEQUENCE 432 AA; 48212 MW; 96FC81E6F0EED01B CRC64;

Query Match 78.9%; Score 1504.5; DB 4; Length 432;  
 Best Local Similarity 80.9%; Pred. No. 1e-123;  
 Matches 292; Conservative 0; Mismatches 0; Indels 69; Gaps 1;

QY 1 MYDNLRGKSGGQYVYEMVGPSPOTLNLIVDTGSSNFVGAAPHEFLHRYQROLSTYR 60  
 Db 63 MYDNLRGKSGGQYVYEMVGPSPOTLNLIVDTGSSNFVGAAPHEFLHRYQROLSTYR 122  
 QY 61 DLKRGVVPYTOGKMEGELGTDLVSIHPGPNVTVRANIAITLSDKPFINSNMEGITL 120  
 Db 123 DLKRGVVPYTOGKMEGELGTDLVSIHPGPNVTVRANIAITLSDKPFINSNMEGITL 145  
 QY 121 AVAELARDDSLPEFFDSLVKQTHVPLNLSLQCGAGFPLNQEVLASVGSMTIGIDH 180  
 Db 146 -----LCGAGFPLNQEVLASVGSMTIGIDH 173  
 QY 181 SLVGTSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDKSIYDSGTTNLRPKYF 240  
 Db 174 SLVGTSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDKSIYDSGTTNLRPKYF 233  
 QY 241 EAAVKSIAASTETKPFDPGFWLGEOLVCMQAGTTPMNIFFVLSLYLMGEVYTNOSFRITIL 300  
 Db 234 EAAVKSIAASTETKPFDPGFWLGEOLVCMQAGTTPMNIFFVLSLYLMGEVYTNOSFRITIL 293  
 QY 301 POQYLRPEVEDVATSDODCKYKFAISQSTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 360

Db 294 POQYLRPEVEDVATSDODCKYKFAISQSTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 353  
 QY 361 H 361  
 Db 354 H 354

## RESULT 5

Q9H2V8 PRELIMINARY; PRT; 439 AA.  
 AC Q9H2V8;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE CDAL3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PHOCHROMOCYTOMA;  
 RA Li Y., Huang Q., Peng, Y., Song H., Yu Y., Xu S., Ren S., Chen Z.,  
 Han Z.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF212252; AAC41783.1; -.  
 DR HSSP; P00797; ZREN.  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 3.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 SQ SEQUENCE 439 AA; 46275 MW; 02EC0E050F11602 CRC64;

Query Match 58.0%; Score 1106; DB 4; Length 439;  
 Best Local Similarity 56.4%; Pred. No. 1e-88;  
 Matches 202; Conservative 60; Mismatches 92; Indels 4; Gaps 2;

QY 1 MYDNLRGKSGGQYVYEMVGPSPOTLNLIVDTGSSNFVGAAPHEFLHRYQROLSTYR 60  
 Db 1 MYDNLQSDSGRGYILELIGTTPQKLIQIVDTGSSNFVAGTTPHSHYDFYDTERSSYR 60  
 QY 61 DLKRGVVPYTOGKMEGELGTDLVSIHPGPNVTVRANIAITLSDKPFINSNMEGITL 120  
 Db 61 SKGPDVYVYKTOGSMWGFVEDLVITPKGNISFLVNIATIFESNFFLPGLKWNGLIGL 120  
 QY 121 AVAELARDDSLPEFFDSLVKQTHVPLNLSLQCGAGFPLNQEVLASVGSMTIGIDH 180  
 Db 121 AVATLAPSSLSLEFFDSLVTOQANINPVFSMQCGAGLFPVAGS--GTNGGSLVIGIEP 177  
 QY 181 SLVGTSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDKSIYDSGTTNLRPKYF 240  
 Db 178 SLVGTSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDKSIYDSGTTNLRPKYF 237  
 QY 241 EAAVKSIAASTETKPFDPGFWLGEOLVCMQAGTTPMNIFFVLSLYLMGEVYTNOSFRITIL 300  
 Db 238 DAVEAIVARASLIPERFDSFWGSQLACWTNSPTSPKISITILRDENSSRSRITIL 297  
 QY 301 POQYLRPEVEDVATSDODCKYKFAISQSTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 358  
 Db 298 POLYIOPMMGAGLNY-ECYRFGISPTNALVIGATVMEGFYVFDRAKRRIQFAVSACHV 354

RESULT 6  
 Q9JL18 PRELIMINARY; PRT; 514 AA.  
 AC Q9JL18;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ASPARTYL PROTEASE 1.  
 GN BACE2.





RP SEQUENCE FROM N.A.  
RA Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;  
RT "Cloning of a gene from chromosome 21 Down Region encoding a potential  
RT transmembrane protease."; to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF051150; AAD45964.1; -  
DR MEROPS: A01.041; -  
DR InterPro: IPR001969; Asp.protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; Asp. 2.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
FT NON TER  
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;  
Query Match 28.7%; Score 547.5; DB 11; Length 255;  
Best Local Similarity 56.7%; Pred. No. 5.2e-40;  
Matches 97; Conservative 31; Mismatches 42; Indels 1; Gaps 1;  
QY 188 WYPTIRREYVYIVRVINGODLKMCKEYNDKSIYDSCGTTNLRPKVFEAAVKSI 247  
DB 1 WYPTIRREYVYIVRVINGODLKMCKEYNDKSIYDSCGTTNLRPKVFEAAVKSI 60  
QY 248 KAASSTKFPDGFVGLVQVQAGTTPNIPVSYLMGSEVTNOSRITLIPQOYLRP 307  
DB 61 ARTSLPEFSDFWTCGAOLACTNSETPWAPYPPKISYLRDNASKRITLIPQOYLRP 120  
QY 308 VEDVATSDDCYKFAISOSTGTVMGAVIMEGFYVDFRARRIGFAVSAC 358  
DB 121 MMGAGENTY-ECIRFGISSSTNALVIGATYMEGFYVDFRARRIGFAVSAC 170  
RESULT 12  
QYKPK6 PRELIMINARY; PRT; 391 AA.  
AC 09VKP6; (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 19, Last annotation update)  
DE CG17134 PROTEIN.  
GN CG17134.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanaides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Morten J.-H., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-R.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abt J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burlis K.C., Busam D.A., Butler H., Cadenot L.B., Davies P.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davis S.M.,  
RA Dodson K., Doup L.E., Doming M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dudin K.T., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glöck A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwan C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattéi B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Szyrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhao W., Zhou X., Zhou Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC EMBL: AE003630; AAF53016.1; -  
CC HSSP: P00794; 4CMS.  
DR MEROPS: A01.0PW; -  
DR FLYbase: FBgn0032304; CG17134.  
DR InterPro: IPR001969; Asp.protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF00026; Asp. 1.  
DR PRINTS: PR00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
DR ASpartyl protease; Hydrolase.  
SQ SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;  
Query Match 17.4%; Score 332.5; DB 5; Length 391;  
Best Local Similarity 29.1%; Pred. No. 7.5e-21;  
Matches 105; Conservative 61; Mismatches 146; Indels 49; Gaps 13;  
QY 3 DNLKSGGGYVEMTVGSSPPTNLILVDGSSNFAVGAAPHF-----LHYRQRLS 56  
DB 66 ENLHSMNNEYGYVIALGTPEGRNILEFTDGSANLWPSASCPASNTACQRHNKYSSAS 125  
QY 57 STYRDLKRGVYVPTQGWEGELGDLVSIPIHGNVYVRANIIATSEDFGINSWMEG 116  
DB 126 SYTVANGGEFAIEVGTGSLSGFLSNDVITTA-GISTONOTGEMLSRGTPVD-AFAG 183  
QY 117 ILGLAVETARPDDSLPEFPDILVKQTHVPN-LFSLQCGAGFLPNOSSEVLASVGSMTI 175  
DB 184 ILGLAFSAIA--VDGVTPEPDMISQGLDEPVISFLYLRKG-----TAVNGGLIL 233  
QY 176 GGIDSLVYSGSLWTPPIREMYEVIYRVELNGODLKMCKEYNDKSIYDSCGTTNLR 235  
DB 234 GGIDSLVYSGSLWTPPIREMYEVIYRVELNGODLKMCKEYNDKSIYDSCGTTNLR 287  
QY 236 PKKVFPAVKSIIKASSTKFPDGFVGLVQVQAGTTPNIPVSYLMGSEVTNOS 294  
DB 288 PLAAVKKIRKOLGALDND-----GEAFVRCGRVSS-----LPKYNLIMGTV--- 329  
QY 295 FRITLIPQOYLRVEDVATSDDCYK-FAISOSTGTVMGAVIMEGFYVDFRARRIGF 353  
DB 330 -FTLAPRDIY--VKVTONGQTCMSAFYMGSLFWILGDVIFGKFTYVDPDKMERIGF 385  
QY 354 A 354  
DB 386 A 386  
RESULT 13  
QYKPK6 PRELIMINARY; PRT; 354 AA.  
AC 09GXY7; (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMblrel. 19, Last annotation update)  
DE HEME-BINDING ASPARTIC PROTEINASE (FRAGMENT).  
OS Boophilus microplus (Cattle tick).  
OC Eukaryota; Metazoa; Arthropoda; Arachnida; Acari;

OC Parasitiformes; Ixodidae; Ixodidae; Boophilus.  
 OX NCBI\_TaxID=6941;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PORTO ALEGRE; TISSUE-Ovary;  
 RA Sordane M.H.F., Logullo C., Zingali R.B., Paiva-Silva G.O.,  
 RA Juliano L., Oliveira P.L.;  
 RT "A heme-binding aspartic proteinase from the eggs of the hard tick  
 RT Boophilus microplus."  
 RL J. Biol. Chem. 0:0-0(2000).  
 DR EMBL; AF286865; AAC00933.1; -  
 DR HSSP; P00797; 2REN.  
 DR MEROPS; A01.054; -  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp. 1.  
 DR PRINTS; PR00792; PEPsin.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 354 AA; 38512 MW; B50D7C638CF27091 CRC64;  
 Query Match 17.4%; Score 332; DB 5; Length 354;  
 Best Local Similarity 26.2%; Pred. No. 7.2e-21;  
 Matches 101; Conservative 74; Mismatches 135; Indels 76; Gaps 15;  
 QY 2 VDNLRGK-----SGGGYVEMTVSGSPQTLNIVDTGSSNFAVGAAPHFL- 47  
 Db 13 VTEIRGALDDPIPIITNTNNMQFGIITIGTPPOSFKILMTGSSNFW-----PSIN 67  
 QY 48 -----HRYQROLSTYRDLRKGVVYPTQGMKEGLGTLVSIPIHGPNTVRANI 98  
 Db 68 CDOSMACROHAKYDSSKSTFTKSGRYIRIRYSGVVRGITSIDNVG--GPATVQYKF 125  
 QY 99 AATSEKPFINGSNMGLGLAYAEIARPDSELPFESLVKQTHVPN-LSLQICGAG 157  
 Db 126 AEMDHSDGKIFRNAKKGDTGLFLAFPSISQ--NQDLPLFAMKQGVVRAVPSLYL--SK 181  
 QY 158 FPLNQSSEVLASVGSMLIGIDHSLYTGSLMYTPIRREWYEIVYIVRVLNQGDKM-DC 216  
 Db 182 QPSEQN-----GGEIFGGINAQRYTGAIHYVPSQAAMHQMVDNINVOGTTLCVGC 235  
 QY 217 KEVNVKSTVDSTNTLRLPKRVFEAAVKSIAASTEKFPDGFVLEQVLCWQAGTTPW 276  
 Db 236 -----PTVVDSTGTSFLSGP---SADVETLNRVIGATKTAGY-----FEVNCATI 277  
 QY 277 NIEPVISLYLMEG---VTNQSFRTITLPOQYLRPVEDVATSDDCYKFAISQSSGT--- 330  
 Db 278 SSILPITFNLNKSFPLOGEATIRI-----PLTTGEGCFTRISESASGTMLM 327  
 QY 331 VMGAVIMEGFYVFDRAKRIRGFASV 356  
 Db 328 ILGAVFTQTYTVFDRANRVGFATA 353  
 RESULT 14  
 O9DEC2 PRELIMINARY; PRT; 384 AA.  
 AC O9DEC2  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PEPsinogen A.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasa S.;  
 RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog  
 RT Rana catesbeiana."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDIJ databases  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

DR EMBL; AB045380; BAB20798.1; -  
 DR HSSP; P00790; 1PSN.  
 DR MEROPS; A01.001; -  
 DR InterPro; IPR001969; Asp-protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp. 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP-PROTEASE; 1.  
 DR Aspartyl protease; Hydrolase.  
 SQ SEQUENCE 384 AA; 41888 MW; 729DD2E7953D9072 CRC64;  
 Query Match 16.2%; Score 308; DB 13; Length 384;  
 Best Local Similarity 27.5%; Pred. No. 1e-18;  
 Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;  
 QY 13 YVEMTVSGSPQTLNIVDTGSSNFAVGAAPHFL-----HRYQROLSTYRDLRK 65  
 Db 72 YGTISIGTPPEFTVIDTDSANLMV---PSYCSSQACSNHNFNQSSSTFOATNTP 128  
 QY 66 VVPTQGMKEGLGTLVSIPIHGPNTVRANIAITESDK-FTINGSNMGLGLAYAE 124  
 Db 129 VSIQGTSMGSEGLGIDYDLQV---GNIOISNQMGISESEGFYSFPDGIIGLAFPS 185  
 QY 125 IARPDSELPFESLVKQTHVP-NLFSLQICGAGPFLNQSSEVLASVGSMLIGIDHSL 183  
 Db 186 IA--SSQATIPVDNMWSSGGLIPQNLFSYLSSDG-----QTSYVLEGGVDSNY 233  
 QY 184 TGLMTYTRREWYEIVYIVRVLNQGDKL--KMDCKEYNKDSIVDSQTNLRPKKVF 241  
 Db 224 SGLSNMVPLEIETWQILDLBVSINGVIACSQSC-----QALVDGTSLMGPSPPI- 286  
 QY 242 AAVKSIAASTEKFPDGFVLEQVLCWQAGTTPPNIPIVTSLYLMEGVTNQSFRITL 300  
 Db 287 ANIQVITASQDSN-----GQYVINCNNISNMPITVF-----TIN 321  
 QY 301 PQQY-LRPVEDVATSDDCYK-FAISQSSGT---VMGAVIMEGFYVFDRAKRIRGFA 354  
 Db 322 GVQYPLSPSAVVRONQCCSSGFQAMNLPITNSGDLMLIGVFIROYTVYDRANNVAYA 381  
 RESULT 15  
 O9DEC4 PRELIMINARY; PRT; 385 AA.  
 AC O9DEC4  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PEPsinogen A.  
 OS Rana catesbeiana (Bull frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.  
 OX NCBI\_TaxID=8400;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasa S.;  
 RT "Molecular Cloning of pepsinogens in Adult Xenopus laevis and Bullfrog  
 RT Rana catesbeiana."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDIJ databases  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 DR EMBL; AB045376; BAB20092.1; -  
 DR HSSP; P00794; 4CMS.  
 DR MEROPS; A01.001; -  
 DR InterPro; IPR001969; Asp-protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp. 1.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP-PROTEASE; 1.  
 DR Aspartyl protease; Hydrolase.  
 SQ SEQUENCE 385 AA; 41702 MW; 5DC8914FC9CEC603 CRC64;  
 Query Match 16.2%; Score 308; DB 13; Length 385;  
 Best Local Similarity 27.8%; Pred. No. 1e-18;





PT Alzheimer's disease -  
XX Claim 10; Fig 3B; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
PS beta-amyloid precursor protein to produce beta-amyloid peptide. This  
XX enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme fragment.  
XX  
XX Sequence 415 AA:  
S0  
Query Match 100.0%; Score 1907; DB 21; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.2e-191;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MYDNLGRKSGGQYVEMTVGSPPTLNLIVDTGSSNFVGAAPHFLHRRYOROLSTYR 60  
DB 18 MYDNLGRKSGGQYVEMTVGSPPTLNLIVDTGSSNFVGAAPHFLHRRYOROLSTYR 77  
OY 61 DLKRGVYPTQGWKGEGLGTDLVSIIPHGPNTVRANTIAITESDKFFINGSMNEGILGL 120  
DB 78 DLKRGVYPTQGWKGEGLGTDLVSIIPHGPNTVRANTIAITESDKFFINGSMNEGILGL 137  
OY 121 AYAETIARPDSDLPEPFDLSLVKOTHPNLFSLDLCGAGPFLNOSSEVLASVGGSMITGGIDH 180  
DB 138 AYAETIARPDSDLPEPFDLSLVKOTHPNLFSLDLCGAGPFLNOSSEVLASVGGSMITGGIDH 197  
OY 181 SLYTGSLWTPPIRREMYEVIIVRVEINGODLKMCKEKNYKSIYDSGTTMLRLPKKVF 240  
DB 198 SLYTGSLWTPPIRREMYEVIIVRVEINGODLKMCKEKNYKSIYDSGTTMLRLPKKVF 257  
OY 241 EAAVKSIRKASSTERKPPDGFMLGEOLVCMQAGTTPWNIFFVISLYLMGEVTNOSFRITTL 300  
DB 256 EAAVKSIRKASSTERKPPDGFMLGEOLVCMQAGTTPWNIFFVISLYLMGEVTNOSFRITTL 317  
OY 301 POQYLRPVEDVATSDDDCKKFAISOSSTGTVMGAVIMEFYVVEPRARRRIGFAVSACHV 360  
DB 318 POQYLRPVEDVATSDDDCKKFAISOSSTGTVMGAVIMEFYVVEPRARRRIGFAVSACHV 377  
OY 361 H 361  
DB 378 H 378  
RESULT 2  
AAV88438  
ID AAV88438 standard; Protein; 453 AA.  
XX  
AC AAV88438;  
XX  
DF 03-AUG-2000 (first entry)  
XX  
DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.  
XX  
KW Aspartyl protease; aspartase; amyloid precursor protein; App; Asp 2;  
KW Alzheimer's disease; beta secretase site.  
XX  
OS Homo sapiens.  
XX  
PN MO200017369-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 23-SEP-1999; 99MO-US20881.  
XX  
PR 24-SEP-1998; 98US-0101594.  
XX

PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
XX Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;  
PI  
XX WPI: 2000-303209/26.  
DR  
XX N-PSDB: AAA15688.  
XX  
PT New enzyme designated human aspartase useful in research into  
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at  
PT the beta secretase site to produce amyloid beta peptide -  
XX  
XX Example 10; Page 169-172; 183pp; English.  
PS  
XX This sequence represents a modified human aspartyl protease 2 (Asp2)  
XX amino acid sequence. Asp2 encoded by this sequence has the C-terminal  
XX transmembrane domain deleted. The invention relates to a protease  
XX (e.g. Asp2) capable of cleaving the beta secretase site of amyloid  
XX precursor protein (App). The protease contains a sequence encoding the  
XX amino acid sequence DTG and a sequence encoding DSG or DTG separated by  
XX 100-300 amino acids. When mutated the App gene causes an autosomal  
XX dominant form of Alzheimer's disease. App localises to the cell surface  
XX membrane and have a single C-terminal transmembrane domain. Proteolytic  
XX processing of App produces the amyloid beta protein, which is possibly  
XX very important in Alzheimer's disease. The invention includes a  
XX nucleotide sequence encoding the protease, a vector containing the  
XX screening for inhibitors of beta secretase activity are also given in the  
XX invention. The human aspartase protein and nucleotide sequences and the  
XX methods for identifying inhibitors of the protease, are useful in the  
XX treatment of and research in to Alzheimer's disease.  
S0  
Sequence 453 AA:  
Query Match 100.0%; Score 1907; DB 21; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.3e-191;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 MYDNLGRKSGGQYVEMTVGSPPTLNLIVDTGSSNFVGAAPHFLHRRYOROLSTYR 60  
DB 63 MYDNLGRKSGGQYVEMTVGSPPTLNLIVDTGSSNFVGAAPHFLHRRYOROLSTYR 122  
OY 61 DLKRGVYPTQGWKGEGLGTDLVSIIPHGPNTVRANTIAITESDKFFINGSMNEGILGL 120  
DB 123 DLKRGVYPTQGWKGEGLGTDLVSIIPHGPNTVRANTIAITESDKFFINGSMNEGILGL 182  
OY 121 AYAETIARPDSDLPEPFDLSLVKOTHPNLFSLDLCGAGPFLNOSSEVLASVGGSMITGGIDH 180  
DB 183 AYAETIARPDSDLPEPFDLSLVKOTHPNLFSLDLCGAGPFLNOSSEVLASVGGSMITGGIDH 242  
OY 181 SLYTGSLWTPPIRREMYEVIIVRVEINGODLKMCKEKNYKSIYDSGTTMLRLPKKVF 240  
DB 243 SLYTGSLWTPPIRREMYEVIIVRVEINGODLKMCKEKNYKSIYDSGTTMLRLPKKVF 302  
OY 241 EAAVKSIRKASSTERKPPDGFMLGEOLVCMQAGTTPWNIFFVISLYLMGEVTNOSFRITTL 300  
DB 303 EAAVKSIRKASSTERKPPDGFMLGEOLVCMQAGTTPWNIFFVISLYLMGEVTNOSFRITTL 362  
OY 301 POQYLRPVEDVATSDDDCKKFAISOSSTGTVMGAVIMEFYVVEPRARRRIGFAVSACHV 360  
DB 363 POQYLRPVEDVATSDDDCKKFAISOSSTGTVMGAVIMEFYVVEPRARRRIGFAVSACHV 422  
OY 361 H 361  
DB 423 H 423  
RESULT 3  
AAE10642  
ID AAE10642 standard; Protein; 453 AA.  
XX  
AC AAE10642;  
XX  
DT 10-DEC-2001 (first entry)



```

XX DE Human-Asp 2(a) protein lacking transmembrane domain.
XX
XX KW Human; aspartyl protease 2a; Asp2a; amyloid precursor protein; APP;
XX KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
XX KW amyloid plaque; neuronal loss; proteolytic; neurotrophic; neuroprotective.
XX OS Homo sapiens.
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 214 /note= "Encoded by CAC"
XX
XX PN GB2357767-A.
XX
XX PD 04-JUL-2001.
XX
XX PE 22-SEP-2000; 2000GB-0023315.
XX
XX PR 23-SEP-1999; 99US-0155493.
XX PR 23-SEP-1999; 99US-0404133.
XX PR 23-SEP-1999; 99WO-US20881.
XX PR 13-OCT-1999; 99US-0416901.
XX PR 06-DEC-1999; 99US-0169232.
XX
XX PA (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX PI Bienkowski MJ, Gurney M;
XX
XX DR WPI: 2001-444208/48.
XX DR N-PSDB; AAD17878.
XX
XX PT Polypeptide comprising fragments of human aspartyl protease with
XX PT amyloid precursor protein processing activity and alpha-secretase
XX PT activity, for identifying modulators useful in treating Alzheimer's
XX PT disease.
XX
XX PS Example 10; Fig 11; 187P; English.
XX
XX CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified
XX CC Asp1 proteins which lack transmembrane domain or amino terminal
XX CC domain or cytoplasmic domain and retains alpha-secretase activity
XX CC and amyloid protein precursor (APP) processing activity. The proteins
XX CC of the invention are useful for assaying hu-Asp1 alpha-secretase
XX CC activity, which in turn is useful for identifying modulators of
XX CC hu-Asp1 alpha-secretase activity, where modulators that increase
XX CC hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's
XX CC disease (AD) which causes progressive dementia with consequent
XX CC formation of amyloid plaques, neurofibrillary tangles, gliosis and
XX CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying
XX CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with
XX CC the substrate under acidic conditions and determining the level of
XX CC hu-Asp1 proteolytic activity. The present sequence is human Asp 2(a)
XX CC protein lacking a transmembrane (TM) domain. This sequence is generated
XX CC by the deletion of the C-terminal TM domain of human Asp 2(a) protein.
XX
XX SO Sequence 453 AA;

```

```

Db 183 AYAETIARPDSDLEFFPSLVKQTHVNFSLQCGAGPFLNDSVLAISVGSMTIGIDH 242
QY 181 SLYTGSMTWTPPIRREMYEVIIVRVEINGQDLKMDCKEYNDKSIYDSGTTNLRPKRVF 240
Db 243 SLYTGSMTWTPPIRREMYEVIIVRVEINGQDLKMDCKEYNDKSIYDSGTTNLRPKRVF 302
QY 241 EAAVKSIRKASSTKPPDSFWMLEQVLCWQACTTWMNIFPVLSYLKMGVNTQSRITIL 300
Db 303 EAAVKSIRKASSTKPPDSFWMLEQVLCWQACTTWMNIFPVLSYLKMGVNTQSRITIL 362
QY 301 PQQYLRPEVDVATSDDCYKFAISOSSTGTVMGAVIMEGFYVFPARRRIGFAVSACHV 360
Db 363 PQQYLRPEVDVATSDDCYKFAISOSSTGTVMGAVIMEGFYVFPARRRIGFAVSACHV 422
QY 361 H 361
Db 423 H 423

```

RESULT 4  
AAE06872  
ID AAE06872 standard; Protein; 453 AA.  
AC AAE06872;  
AC  
DT 23-OCT-2001 (first entry)  
DT  
XX  
DE Human-Asp2(a) deltaTM protein.  
XX  
KW Human; aspartyl protease 2a; Asp 2a; beta-amyloid precursor protein; APP;  
KW beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;  
KW neurofibrillary tangle; neuronal loss; amyloid-beta peptide; neurotrophic;  
KW neuroprotective; antisense therapy; Asp2(a) deltaTM protein;  
KW gene therapy.  
XX  
XX OS Homo sapiens.  
XX OS Synthetic.  
XX  
XX FH Key Location/Qualifiers  
XX FT Misc-difference 214 /note= "Encoded by CAC"  
XX  
XX PN WO200150829-A2.  
XX  
XX PD 19-JUL-2001.  
XX  
XX PR 09-MAY-2001; 2001WO-IB00799.  
XX PR 09-MAY-2001; 2001WO-IB00799.  
XX PR 09-MAY-2001; 2001WO-IB00799.  
XX  
XX PA (BIEN/) BIENKOWSKI M J.  
XX PA (GURN/) GURNEY M E.  
XX PA (HEIN/) HEINRIKSON R L.  
XX PA (PARO/) PARODI L A.  
XX PA (YANR/) YAN R.  
XX  
XX PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
XX  
XX DR WPI: 2001-483072/52.  
XX DR N-PSDB; AAD13034.  
XX  
XX PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
XX PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
XX PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
XX PT activity.  
XX  
XX PS Claim 149; Fig 11; 185P; English.  
XX  
XX CC The invention relates to human aspartyl proteases (hu-Asp), beta-amyloid  
XX CC precursor protein (APP) isoforms and their corresponding DNA molecules.  
XX CC Human aspartyl proteases can act as beta-secretase proteases useful for  
XX CC treating Alzheimer's disease. APP isoforms are useful for identifying  
XX CC modulators of amyloid-beta peptide production, for use in designing

therapeutics for the treatment and prevention of Alzheimer's disease, dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis and neuronal loss. APP isoforms are also used in methods for identifying inhibitors and modulators of human Asp2 activity. The invention relates to a method for identifying agents that modulate the activity of human aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used as a means to screen in cellular assays for the inhibitors of beta- and gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in polymerase chain reactions (PCR). The probes are useful for detecting Hu-Asp nucleic acids in in vitro assays and in Northern and Southern blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a) deltam protein which is obtained by the deletion of transmembrane domain at the C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase activity.

Sequence 453 AA:

Query Match 100.0%; Score 1907; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-191;  
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDNLRGKSGGGYVEMTVGSPPTNLILVDTGSSNFVGAAPHPFLHRYQRLSSTYR 60  
 DB 63 MVDNLRGKSGGGYVEMTVGSPPTNLILVDTGSSNFVGAAPHPFLHRYQRLSSTYR 122  
 QY 61 DLKRGVYVPTGKMGEGELGTDLVSIPIHGPNVTVRANIAAITESDKFFINGSMWEGIIGL 120  
 DB 123 DLKRGVYVPTGKMGEGELGTDLVSIPIHGPNVTVRANIAAITESDKFFINGSMWEGIIGL 182  
 QY 121 AYAEIARPDSDLPPFDLSLVKQTHVNPFLSLQCGAGFPINQSEVLASVGGSMIIIGIDH 180  
 DB 183 AYAEIARPDSDLPPFDLSLVKQTHVNPFLSLQCGAGFPINQSEVLASVGGSMIIIGIDH 242  
 QY 181 SLVTGSLMTYPIRREMYEVIIVRVEINQDILKMDCKEYNYDKSIYDSTGTLRLPKKYF 240  
 DB 243 SLVTGSLMTYPIRREMYEVIIVRVEINQDILKMDCKEYNYDKSIYDSTGTLRLPKKYF 302  
 QY 241 EAAVSIKAASTKPEPDGFWLGEOLVCMQAGTTPNINIPVLSILMGSEVTVNSFRITTL 300  
 DB 303 EAAVSIKAASTKPEPDGFWLGEOLVCMQAGTTPNINIPVLSILMGSEVTVNSFRITTL 362  
 QY 301 POQYLKRPEDVATSQDDCYKFAISQSGTGTVMGAVIMEGFYVYFDRARRKIGFVAVSACHV 360  
 DB 363 POQYLKRPEDVATSQDDCYKFAISQSGTGTVMGAVIMEGFYVYFDRARRKIGFVAVSACHV 422  
 QY 361 H 361  
 DB 423 H 423

RESULT 5  
 AAU06616  
 ID AAU06616 standard; Protein; 453 AA.  
 AC AAU06616;  
 XX  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human-pro-Asp 2(a) delta TM.  
 XX  
 KW Human; Aspartyl protease; beta-secretase; noctropic; ASP2;  
 KM neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
 KW amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT Protein /label= Signal\_peptide  
 FT Misc-difference 22..453  
 FT /label= Mature\_Human\_pro-Asp\_2(a)\_delta\_TM  
 FT 214

/note="Encoded by CAC"

FT  
 XX  
 PN WC0200149098-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 PF 09-MAY-2001; 2001WO-IB00798.  
 XX  
 PR 09-MAY-2001; 2001WO-IB00798.  
 XX  
 PA (BIEN/) BIENKOWSKI M J.  
 PA (GURNEY/) GURNEY M E.  
 PA (HEIN/) HEINRIKSON R L.  
 PA (PARODI/) PARODI L A.  
 PA (YANR/) YAN R.  
 PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
 XX WPI; 2001-502549/55.  
 DR N-PSDB; AAS11530.  
 XX  
 PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
 PT activity  
 PS Claim 149; Page 160; 185pp; English.  
 XX  
 CC The invention relates to a purified polypeptide comprising a fragment of  
 CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2  
 CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
 CC the fragment retain the beta-secretase activity of the mammalian Asp2  
 CC protein. The invention also details polynucleotides for the Asp  
 CC proteins and vectors expressing them, and a polypeptide (isoform of  
 CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
 CC APP or its fragment containing an APP cleavage site recognizable by a  
 CC mammalian beta-secretase, and further comprising two lysine residues at  
 CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
 CC APP fragment. Also included in the invention are methods of identifying  
 CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
 CC useful for treating Alzheimer's disease. APP is useful in methods for  
 CC identifying inhibitors or modulators of human Asp2 activity and  
 CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
 CC therapeutics for the treatment or prevention of Alzheimer's disease.  
 CC APP comprising the APP-sw beta-secretase peptide sequence (NMDA), which  
 CC is associated with increased levels of Abeta processing is useful in  
 CC assays relating to Alzheimer's research. The expression vector is useful  
 CC for recombinantly expressing APP. Nucleic acids that hybridise to  
 CC APP oligonucleotides are useful as probes or primers. The probes are  
 CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
 CC Northern and Southern blots. The present sequence is Human-pro-  
 CC Asp 2(a) delta TM protein, which lacks the C-terminal transmembrane  
 CC domain.  
 XX  
 XX Sequence 453 AA:

Query Match 100.0%; Score 1907; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-191;  
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDNLRGKSGGGYVEMTVGSPPTNLILVDTGSSNFVGAAPHPFLHRYQRLSSTYR 60  
 DB 63 MVDNLRGKSGGGYVEMTVGSPPTNLILVDTGSSNFVGAAPHPFLHRYQRLSSTYR 122  
 QY 61 DLKRGVYVPTGKMGEGELGTDLVSIPIHGPNVTVRANIAAITESDKFFINGSMWEGIIGL 120  
 DB 123 DLKRGVYVPTGKMGEGELGTDLVSIPIHGPNVTVRANIAAITESDKFFINGSMWEGIIGL 182  
 QY 121 AYAEIARPDSDLPPFDLSLVKQTHVNPFLSLQCGAGFPINQSEVLASVGGSMIIIGIDH 180  
 DB 183 AYAEIARPDSDLPPFDLSLVKQTHVNPFLSLQCGAGFPINQSEVLASVGGSMIIIGIDH 242  
 QY 181 SLVTGSLMTYPIRREMYEVIIVRVEINQDILKMDCKEYNYDKSIYDSTGTLRLPKKYF 240

Db 243 SLVTSGLWTPPIRREWEYEVIIIVREINGODLKMCKEYNDKSIYDSGTNLRPKVF 302  
QY 241 EAAVKSIRKASSTKEKFPDGFMLGEOIYVCMOAGTTPMNTFPIVSIYLKGEVINOFSRITIL 300  
Db 303 EAAVKSIRKASSTKEKFPDGFMLGEOIYVCMOAGTTPMNTFPIVSIYLKGEVINOFSRITIL 362  
QY 301 POQYLRPEVDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVEDRARRKIRIGFAVSACHV 360  
Db 363 POQYLRPEVDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVEDRARRKIRIGFAVSACHV 422  
QY 361 H 361  
Db 423 H 423

RESULT 6  
AAU07215  
ID AAU07215 standard; Protein: 453 AA.  
XX  
AC AAU07215:  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human aspartyl protease 2a deltatm (HuAsp-2adeltatm).  
XX  
KM Human: aspartyl protease 1; Asp-1; noctropic; neuroprotective;  
KM aspartyl protease 2; Asp2; amyloid protein precursor; APP;  
XX beta-secretase; Alzheimer's disease; HuAsp-2adeltatm.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT Protein /note="Signal peptide"  
FT 22..453  
FT /note="Mature human aspartyl protease 2a deltatm"  
XX  
PN MO200149097-A2.  
XX  
PD 12-JUL-2001.  
XX  
PE 09-MAY-2001; 2001MO-IB00797.  
XX  
PR 09-MAY-2001; 2001MO-IB00797.  
XX  
PA (BIEN/) BIENKOWSKI M J.  
PA (GURN/) GURNEY M E.  
PA (HEIN/) HEINRIKSON R L.  
PA (PARO/) PARODI L A.  
PA (YANR/) YAN R.  
XX  
PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX WPI: 2001-502548/55.  
XX DR N-PSDB: AAS11715.  
XX  
PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity  
XX  
XX Claim 149; Fig 11; 185pp; English.  
XX  
XX The invention relates to a novel purified polypeptide comprising a  
XX fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
XX Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
XX and the fragment retain the beta-secretase activity of the mammalian Asp2  
XX protein. Also included is an isoform of amyloid protein precursor (APP)  
XX comprising the amino acid sequence of a APP or its fragment containing  
XX an APP cleavage site recognizable by a mammalian beta-secretase, and  
XX further comprising two lysine residues at the carboxyl terminus of the  
XX amino acid sequence of the mammalian APP or APP fragment. The

CC polypeptides are used for assaying for modulators of beta-secretase  
CC activity; identifying agents that inhibit the APP processing activity  
CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that  
CC modulate the activity of Asp2; and for reducing cellular production of  
CC amyloid beta (Abeta) from APP. Agents identified by the above methods  
CC are useful for treating Alzheimer's disease, and for identifying  
CC modulators of amyloid-beta (Abeta) peptide production, for use in  
CC designing therapeutics for the treatment or prevention of Alzheimer's  
CC disease. Probes and primers derived from Asp nucleic acid sequences  
CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence represents the amino  
CC acid sequence of human Asp-2a delta TM construct which lacks the amino  
CC transmembrane domain. This construct was used for bacterial expression  
CC and purification of human Asp2a.  
XX

Seq Sequence 453 AA;  
Query Match 100.0%; Score 1907; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.3e-191;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDNLGKSGGGYVEMIVGSPPTNLILVDTGSSNFVAGAAPHPFLHRYORQLSTYR 60  
Db 63 MVDNLGKSGGGYVEMIVGSPPTNLILVDTGSSNFVAGAAPHPFLHRYORQLSTYR 122  
QY 61 DLKRGYVYPTGKMGEGELDTLVSTPHGPNTVYRANIAITESDKFTNGSNMGGITGL 120  
Db 123 DLKRGYVYPTGKMGEGELDTLVSTPHGPNTVYRANIAITESDKFTNGSNMGGITGL 182  
QY 121 AYAELARPDDSLPEPFDLSLVKQTHVNLFSIOLCGAGFPPLNOSEVLASVGSMTIGTIDH 180  
Db 183 AYAELARPDDSLPEPFDLSLVKQTHVNLFSIOLCGAGFPPLNOSEVLASVGSMTIGTIDH 242  
QY 181 SLVTSGLWTPPIRREWEYEVIIIVREINGODIKMCKEYNDKSIYDSGTNLRPKVF 240  
Db 243 SLVTSGLWTPPIRREWEYEVIIIVREINGODIKMCKEYNDKSIYDSGTNLRPKVF 302  
QY 241 EAAVKSIRKASSTKEKFPDGFMLGEOIYVCMOAGTTPMNTFPIVSIYLKGEVINOFSRITIL 300  
Db 303 EAAVKSIRKASSTKEKFPDGFMLGEOIYVCMOAGTTPMNTFPIVSIYLKGEVINOFSRITIL 362  
QY 301 POQYLRPEVDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVEDRARRKIRIGFAVSACHV 360  
Db 363 POQYLRPEVDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVEDRARRKIRIGFAVSACHV 422  
QY 361 H 361  
Db 423 H 423

RESULT 7  
AAE02594  
ID AAE02594 standard; Protein: 453 AA.  
XX  
AC AAE02594:  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Human-Asp-2(a) delta TM protein.  
XX  
XX Human: alpha-secretase; amyloid precursor protein; APP; therapy;  
XX Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp2a;  
XX beta-secretase; Asp-2a delta TM.  
XX  
XX Homo sapiens.  
XX OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Msc-difference 214 /note="Encoded by CAC"  
FT  
XX  
XX PN MO200123533-A2.  
XX

Query Match	Best Local Similarity	Matches 361;	Score 1907;	DB 22;	Length 453;
100.0%;	100.0%;	0;	Mismatches 0;	Indels 0;	Gaps 0;
1	MDNLRKSGSGGYVETWVSSPQTLNIIIVDTGSSNFVAGAPHPFLHRYQRQLSSTR	60			
63	MDNLRKSGSGGYVETWVSSPQTLNIIIVDTGSSNFVAGAPHPFLHRYQRQLSSTR	122			
61	DLRKGVPYPTQGWKGEGLDLYSTPHGNVYVRANIAITESDKPFINGWEGILGL	120			
123	DLRKGVPYPTQGWKGEGLDLYSTPHGNVYVRANIAITESDKPFINGWEGILGL	182			
121	AVARIARDDSLRPFEDFSLVKOTHPNLFSLQLCGAGFPLNOSSEVLASVGGSMIGGIDH	180			
183	AVARIARDDSLRPFEDFSLVKOTHPNLFSLQLCGAGFPLNOSSEVLASVGGSMIGGIDH	242			
181	SLVYGSIMYTPPIRREMYEVIIVVEINIGDGLKMDCKEYNDKSYDSGTTNLRPKYF	240			
243	SLVYGSIMYTPPIRREMYEVIIVVEINIGDGLKMDCKEYNDKSYDSGTTNLRPKYF	302			
241	EAAYKSIKAASSTKFPDGFMLGEOLOVOMOGATPPMNIIPVLSLYLNGEYTNOSFRITIL	300			
303	EAAYKSIKAASSTKFPDGFMLGEOLOVOMOGATPPMNIIPVLSLYLNGEYTNOSFRITIL	362			
301	POOYLRPEDVATSDDCYKFAISQSSGTGVGAVIMEGFYVDFRARRKIRGFVAVSACHV	360			
363	POOYLRPEDVATSDDCYKFAISQSSGTGVGAVIMEGFYVDFRARRKIRGFVAVSACHV	422			
361	H 361				
423	H 423				

DE	Active enzyme portion of human beta-secretase enzyme.
XX	Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
KW	Amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
KM	Inhibitor; ss.
XX	Homo sapiens.
OS	
PN	WO20047618-A2.
PD	17-Aug-2000.
XX	
PF	10-FEB-2000; 2000WO-US03819.
XX	
PR	10-FEB-1999; 99US-0119571.
XX	
PA	15-JUN-1999; 99US-0139172.
PI	(ELAN) ELAN PHARM INC.
PB	
PT	Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M; Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L; WPI; 2000-533011/48.
PS	
PP	Purified beta-secretase protein used in assays to discover inhibitors which can be used for the treatment of amyloidogenic diseases e.g. Alzheimer's disease -
XX	
PS	Claim 24; Fig 2B; 12pp; English.
XX	
CC	The specification describes a beta-secretase enzyme. The enzyme cleaves beta-amyloid precursor protein to produce beta-amyloid peptide. This enzyme is therefore implicated in the production of amyloid plaque components which accumulate in the brains of individuals afflicted with Alzheimer's disease. Inhibitors of beta-secretase are administered to a mammalian subject e.g. with Alzheimer's disease or Alzheimer's disease-like pathology to test if they maintain or improve cognitive ability or reduce the plaque burden. The compounds are used for the treatment of amyloidogenic diseases e.g. Alzheimer's disease. The present sequence represents the active enzyme portion of human beta-secretase enzyme.
CC	
XX	
SQ	Sequence 456 AA:
Query Match	100.0%; Score 1907; DB 21; Length 456;
Best Local Similarity	100.0%; Pred. No. 1.3e-191;
Matches 361; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MWDLNRKSGQGYYEVTGSSPQTNIIVDGRSSNFAVGAPHPFLHRYRQLSTSYR 60
DB	18 MWDLNRKSGQGYYEVTGSSPQTNIIVDGRSSNFAVGAPHPFLHRYRQLSTSYR 77
QY	61 DLRGVVPYPTGGKEGLGTDLVSLPHGNPVVRANIAITESDKFFINGSMWEGILGL 120
DB	78 DLRGVVPYPTGGKEGLGTDLVSLPHGNPVVRANIAITESDKFFINGSMWEGILGL 137
QY	121 AYAEIARDDSLPEFDFSLVKQTHVPMFLSLQCGAGFLPNQSEVLASVGSMTIGIDH 180
DB	138 AYAEIARDDSLPEFDFSLVKQTHVPMFLSLQCGAGFLPNQSEVLASVGSMTIGIDH 197
QY	181 SLTYGSLMYPTIRRENYEVETIVEINQODLKMDCKEYNVDKSIYDSGTTNLRLPKPY 240
DB	198 SLTYGSLMYPTIRRENYEVETIVEINQODLKMDCKEYNVDKSIYDSGTTNLRLPKPY 257
QY	241 EAAVSISKAASSTEKFPDGFWLGHOLYQWOGATTPMNIFPVISLYLMGEVYNOSFRITLL 300
DB	258 EAAVSISKAASSTEKFPDGFWLGHOLYQWOGATTPMNIFPVISLYLMGEVYNOSFRITLL 317
QY	301 PQOYLRPVEDVATSDDCYKFAISOSSSTGYTMGAVINMEGFVVEDRRARRKRIGFVAVSACHV 360
DB	318 PQOYLRPVEDVATSDDCYKFAISOSSSTGYTMGAVINMEGFVVEDRRARRKRIGFVAVSACHV 377
QY	361 H 361



DR WPI: 2001-444208/48.  
 DR N-PSDB; AAD1/879.  
 XX Polypeptide comprising fragments of human aspartyl protease with  
 PT amyloid precursor protein processing activity and alpha-secretase  
 PT activity, for identifying modulators useful in treating Alzheimer's  
 PT disease -  
 XX  
 PS Example 10; Fig 12; 187pp; English.  
 CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified  
 CC Asp1 proteins which lack transmembrane domain or amino terminal  
 CC domain or cytoplasmic domain and retains alpha-secretase activity  
 CC and amyloid precursor (APP) processing activity. The proteins  
 CC of the invention are useful for assaying hu-Asp1 alpha-secretase  
 CC activity, which in turn is useful for identifying modulators of  
 CC hu-Asp1 alpha-secretase activity, where modulators that increase  
 CC hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's  
 CC disease (AD) which causes progressive dementia with consequent  
 CC formation of amyloid plaques, neurofibrillary tangles, gliosis and  
 CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying  
 CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with  
 CC the substrate under acidic conditions and determining the level of  
 CC hu-Asp1 proteolytic activity. The present sequence is human Asp 2(a)  
 CC protein lacking a transmembrane (TM) domain and containing (his)6  
 CC sequence. This sequence is generated from human Asp 2(a) protein by  
 CC the deletion of its C-terminal TM domain and addition of hexa-histidine  
 CC tag at its C-terminus.  
 CC  
 XX Sequence 459 AA:  
 SQ  
 Query Match 100.0%; Score 1907; DB 22; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-191;  
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 MPDNLKRGSGGYYVMTYSGSPPTLNIIVDTGSSNFVGAAPHPLHRYROLSTYR 60  
 63 MPDNLKRGSGGYYVMTYSGSPPTLNIIVDTGSSNFVGAAPHPLHRYROLSTYR 122  
 61 DLKRGVYVYTGQKMGELGDTLVSPHGPNTYVRANIAITESDKEFFNGSMWESILGL 120  
 123 DLKRGVYVYTGQKMGELGDTLVSPHGPNTYVRANIAITESDKEFFNGSMWESILGL 182  
 121 AYAEIARPDSDLEPPFDSLKYKQTHVNLFSLOCGAGPFLNQEVLASVGGSMITGIDH 180  
 183 AYAEIARPDSDLEPPFDSLKYKQTHVNLFSLOCGAGPFLNQEVLASVGGSMITGIDH 242  
 181 SLVTGSLMTYPIRREMYEVIIVVEINSGDLKMDCKEYNDKSIYDSGTTNRLPKKYF 240  
 243 SLVTGSLMTYPIRREMYEVIIVVEINSGDLKMDCKEYNDKSIYDSGTTNRLPKKYF 302  
 241 EAAVKSITKAASSTREKPPDGVNLCQVLCWQAGTTPNNIPVLSILYMGVETNOSFRITL 300  
 303 EAAVKSITKAASSTREKPPDGVNLCQVLCWQAGTTPNNIPVLSILYMGVETNOSFRITL 362  
 301 POOYLRFVEDVATSDDCYCFATISQSGTGTVMGAVIMEGFYVVEDRARRKRGFAVSACHV 360  
 363 POOYLRFVEDVATSDDCYCFATISQSGTGTVMGAVIMEGFYVVEDRARRKRGFAVSACHV 422  
 361 H 361  
 423 H 423  
 RESULT 11  
 AAE06873  
 ID AAE06873 standard; Protein: 459 AA.  
 XX  
 AC AAE06873;  
 XX  
 DT 23-OCT-2001 (first entry)  
 XX  
 DE Human-Asp2(a) deltaTM (His)6 protein.

XX Human: aspartyl protease 2a; Asp 2a; beta-amyloid precursor protein; APP;  
 KW beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;  
 KW neurofibrillary tangle; neuronal loss; amyloid-beta peptide; neurotrophic;  
 KW neuroprotective; antisense therapy; Asp2(a) deltaTM (His)6 protein;  
 KW gene therapy.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 FH Peptide 22..30  
 FT /label= Signal\_peptide  
 FT Protein 31..459  
 FT /note= "Mature Human-Asp2(a) deltaTM (His)6 protein"  
 FT Region 1..22  
 FT /note= "corresponds to N-terminal Human-Asp2(a) deltaTM  
 (His)6 protein"  
 FT Misc-difference 214  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 254  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 255  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 256  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 257  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 258  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 259  
 FT /note= "Encoded by CAG"  
 FT Misc-difference 259  
 FT /note= "Encoded by CAG"  
 PD WO200150829-A2.  
 XX 19-JUL-2001.  
 PF 09-MAY-2001; 2001WO-IB00799.  
 XX 09-MAY-2001; 2001WO-IB00799.  
 PR 09-MAY-2001; 2001WO-IB00799.  
 PA (BIEN/) BIENKOWSKI M J.  
 PA (GURN/) GURNEY M E.  
 PA (HEIN/) HEINRIKSON R L.  
 PA (PARO/) PARODI L A.  
 PA (YANR/) YAN R.  
 PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
 DR WPI: 2001-483072/52.  
 DR N-PSDB; AAD13035.  
 PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
 PT secretase activity of Asp2 useful for identifying inhibitors of beta  
 PT activity -  
 XX  
 PS Claim 149; Fig 12; 185pp; English.  
 XX  
 CC The invention relates to human aspartyl proteases (Hu-Asp), beta-amyloid  
 CC precursor protein (APP) isoforms and their corresponding DNA molecules.  
 CC Human aspartyl proteases can act as beta-secretase proteases useful for  
 CC treating Alzheimer's disease. APP isoforms are useful for identifying  
 CC modulators of amyloid-beta peptide production, for use in designing  
 CC therapeutics for the treatment and prevention of Alzheimer's disease,  
 CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis  
 CC and neuronal loss. APP isoforms are also used in methods for identifying  
 CC inhibitors and modulators of human Asp2 activity. The invention relates  
 CC to a method for identifying agents that modulate the activity of human  
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
 CC as a means to screen in cellular assays for the inhibitors of beta- and  
 CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
 CC polymerase chain reactions (PCR). The probes are useful for detecting

CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
CC blocks. The present sequence is Human aspartyl protease 2a (Hu-Asp2a)  
CC deltatm (His)6 protein which is obtained by deletion of C-terminal  
CC transmembrane domain and addition of a hexa-Histidine tag at the  
CC C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase activity.  
XX

SO Sequence 459 AA:

Query Match 100.0%; Score 1907; DB 22; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1,4e-191;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDNLRGKSGGGYVEMTGSPPTLNTLVDTGSSNFVGAAPHPEFLHRYQRLSTYR 60  
DB 63 MYDNLRGKSGGGYVEMTGSPPTLNTLVDTGSSNFVGAAPHPEFLHRYQRLSTYR 122  
QY 61 DLKRGVYVYTGKMGELGTDLVSIPIGPNVTYRANIAATTESDKFFINSNNEGILGL 120  
DB 123 DLKRGVYVYTGKMGELGTDLVSIPIGPNVTYRANIAATTESDKFFINSNNEGILGL 182  
QY 121 AYAELARPDDSLPEPFDSLVKQTHVNFSLQCGAGFPLNQSEVLASVGSMTIGIDH 180  
DB 183 AYAELARPDDSLPEPFDSLVKQTHVNFSLQCGAGFPLNQSEVLASVGSMTIGIDH 242  
QY 181 SLYTGSIMWTPIRREMYEVIIVREINGQDLKMDCKEYNKSIYDSGTTNLRPKKVF 240  
DB 243 SLYTGSIMWTPIRREMYEVIIVREINGQDLKMDCKEYNKSIYDSGTTNLRPKKVF 302  
QY 241 EAAVKSIAKASSTEFKPDGFWLGEOLVCMQAGTTPMNIIPVLSILMGVTVNOSFRITL 300  
DB 303 EAAVKSIAKASSTEFKPDGFWLGEOLVCMQAGTTPMNIIPVLSILMGVTVNOSFRITL 362  
QY 301 POQYLRPEVDVATSDDDCKFAISQSSSTGTVMGAVIMEGFYVFPDRARRKIGFAVSACHV 360  
DB 363 POQYLRPEVDVATSDDDCKFAISQSSSTGTVMGAVIMEGFYVFPDRARRKIGFAVSACHV 422  
QY 361 H 361  
DB 423 H 423

RESULT 12  
AAU06617  
ID AAU06617 standard; Protein: 459 AA.

XX AAU06617;

DT 24-OCT-2001 (first entry)

DE Human-pro-Asp 2(a) delta TM (His)6.

XX Human; Aspartyl protease; beta-secretase; nootropic; ASP2;  
KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;  
KW amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM (His)6; mutant; mutain.  
XX

OS Homo sapiens.  
XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /label= signal\_peptide

FT MISC-difference /label= Mature\_Human\_pro\_Asp\_2(a)\_delta\_TM\_(His)6

FT MISC-difference 214 /note= "Encoded by CAC"

FT MISC-difference 454..459 /note= "Encoded by CAGCAGCAGCAGCAGCAG"

FT Region 454..459 /label= His-tag

FT /note= "Nickel binding region to aid purification"

XX W0200149098-A2.

PD 12-JUL-2001.

XX 09-MAY-2001; 2001WO-1B00798.

PR 09-MAY-2001; 2001WO-1B00798.

XX (BIEN/) BIENKOWSKI M J.

XX (GURN/) GURNEY M E.

XX (HEIN/) HEINRIKSON R L.

XX (PARO/) PARODI L A.

XX (YANR/) YAN R.

XX Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;

PI WPI: 2001-502549/55.

DR N-PSDB; AAS11531.

PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity

PS Claim 149; Fig 12; 185pp; English.

XX The invention relates to a purified polypeptide comprising a fragment of  
CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2  
CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
CC the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. The invention also details polynucleotides for the Asp  
CC proteins and vectors expressing them, and a polypeptide (isoform of  
CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
CC APP or its fragment containing an APP cleavage site recognizable by a  
CC mammalian beta-secretase, and further comprising two lysine residues at  
CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
CC APP fragment. Also included in the invention are methods of identifying  
CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
CC useful for treating Alzheimer's disease. APP is useful in methods for  
CC identifying inhibitors or modulators of human Asp2 activity and  
CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
CC therapeutics for the treatment or prevention of Alzheimer's disease.  
CC APP comprising the APP-Sw-beta-secretase peptide sequence (NLDN), which  
CC is associated with increased levels of Abeta processing is useful in  
CC assays relating the Alzheimer's research. The expression vector is useful  
CC for recombinantly expressing APP. Nucleic acids that hybridize to  
CC Asp oligonucleotides are useful as probes or primers. The probes are  
CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence is Human-pro-  
CC Asp 2(a) delta TM (His)6 protein, which lacks the C-terminal  
CC transmembrane domain and has a His tag to aid purification.  
XX

SO Sequence 459 AA:

Query Match 100.0%; Score 1907; DB 22; Length 459;  
Best Local Similarity 100.0%; Pred. No. 1,4e-191;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDNLRGKSGGGYVEMTGSPPTLNTLVDTGSSNFVGAAPHPEFLHRYQRLSTYR 60  
DB 63 MYDNLRGKSGGGYVEMTGSPPTLNTLVDTGSSNFVGAAPHPEFLHRYQRLSTYR 122  
QY 61 DLKRGVYVYTGKMGELGTDLVSIPIGPNVTYRANIAATTESDKFFINSNNEGILGL 120  
DB 123 DLKRGVYVYTGKMGELGTDLVSIPIGPNVTYRANIAATTESDKFFINSNNEGILGL 182  
QY 121 AYAELARPDDSLPEPFDSLVKQTHVNFSLQCGAGFPLNQSEVLASVGSMTIGIDH 180  
DB 183 AYAELARPDDSLPEPFDSLVKQTHVNFSLQCGAGFPLNQSEVLASVGSMTIGIDH 242  
QY 181 SLYTGSIMWTPIRREMYEVIIVREINGQDLKMDCKEYNKSIYDSGTTNLRPKKVF 240  
DB 243 SLYTGSIMWTPIRREMYEVIIVREINGQDLKMDCKEYNKSIYDSGTTNLRPKKVF 302  
QY 241 EAAVKSIAKASSTEFKPDGFWLGEOLVCMQAGTTPMNIIPVLSILMGVTVNOSFRITL 300

```

Db      303 EAAVKSIRKAASSTKRPDGFGLGEOIYVWQAGTTPWNIPEYISLYLMGEVTNOSFRITIL 362
OY      301 POQYLRPVEDVATSDDDCYKFAISOSSTGTVMGAVIMGEFYVVDRAKRRIGFAVSACHV 360
Db      363 POQYLRPVEDVATSDDDCYKFAISOSSTGTVMGAVIMGEFYVVDRAKRRIGFAVSACHV 422
OY      361 H 361
Db      423 H 423

RESULT 13
AA007216
ID      AA007216 standard; Protein; 459 AA.
AC      AA007216;
XX      24-OCT-2001 (first entry)
DE      Human aspartyl protease 2a deltatm (His)6.
XX      Human; aspartyl protease 1; Asp-1; neurotrophic; neuroprotective;
KW      aspartyl protease 2; Asp2; amyloid protein precursor; App;
KM      beta-secretase; Alzheimer's disease; HuAsp-2adeltatm (His)6.
XX      Homo sapiens.
OS
FH      Key
FH      Peptide
FT      1..21
FT      /note= "Signal peptide"
FT      22..459
FT      /note= "Mature human aspartyl protease 2a deltatm (His)6"
FT      Misc-difference 454
FT      /note= "Encoded by cag"
FT      Misc-difference 455
FT      /note= "Encoded by cag"
FT      Misc-difference 456
FT      /note= "Encoded by cag"
FT      Misc-difference 457
FT      /note= "Encoded by cag"
FT      Misc-difference 458
FT      /note= "Encoded by cag"
FT      Misc-difference 459
FT      /note= "Encoded by cag"
FT      /note= "Encoded by cag"
PN      WO200149097-A2.
PD      12-JUL-2001.
PF      09-MAY-2001; 2001WO-1B00797.
XX      09-MAY-2001; 2001WO-1B00797.
PR      09-MAY-2001; 2001WO-1B00797.
XX      (BIEN/) BIENKOWSKI M J.
PA      (GURNEY/) GURNEY M E.
PA      (HEIN/) HEINRIKSON R L.
PA      (PARO/) PARODI L A.
PA      (YANR/) YAN R.
XX      Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;
PI      WPI: 2001-502548/55.
XX      DR      N-PSDB; AAS11716.
XX      Novel purified polypeptide comprising fragment of mammalian aspartyl
PT      protease 2, lacking Asp2 transmembrane domain and retaining beta
PT      secretase activity of Asp2 useful for identifying inhibitors of Asp2
PT      activity
XX      Claim 149; Fig 12; 185pp; English.
XX      CC      The invention relates to a novel purified polypeptide comprising a

```

```

CC      fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the
CC      Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide
CC      and the fragment retain the beta-secretase activity of the mammalian Asp2
CC      protein. Also included is an isoform of amyloid protein precursor (APP)
CC      comprising the amino acid sequence of a APP or its fragment containing
CC      an APP cleavage site recognisable by a mammalian beta-secretase, and
CC      further comprising two lysine residues at the carboxyl terminus of the
CC      amino acid sequence of the mammalian APP or APP fragment. The
CC      polypeptides are used for assaying for modulators of beta-secretase
CC      activity; identifying agents that inhibit the APP processing activity
CC      of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that
CC      modulate the activity of Asp2; and for reducing cellular production of
CC      amyloid beta (Abeta) from APP. Agents identified by the above methods
CC      are useful for treating Alzheimer's disease; and for identifying
CC      modulators of amyloid-beta (beta) peptide production, for use in
CC      designing therapeutics for the treatment or prevention of Alzheimer's
CC      disease. Probes and primers derived from Asp nucleic acid sequences
CC      are useful for detecting Hu-Asp nucleic acids in in vitro assays and in
CC      Northern and Southern blots. The present sequence represents the amino
CC      acid sequence of human Asp-2a delta TM (His)6 construct which has
CC      a 6 histidine tag and lacks the transmembrane domain. This construct was
CC      used for expression and purification of human Asp2a in insect cells.
XX
SQ      Sequence 459 AA:

```

```

Query Match          100.0%; Score 1907; DB 22; Length 459;
Best Local Similarity 100.0%; Pred. No. 1,4e-191;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 MYDNLRGKSGGQYVEMVGSPPOTLNLIVDTGSSNFVGAAPHFLLRRYOROLSTSYR 60
Db      63 MYDNLRGKSGGQYVEMVGSPPOTLNLIVDTGSSNFVGAAPHFLLRRYOROLSTSYR 122
OY      61 DLRKGVYVPTGKWEGLGDTLVSIHPGPNVTYANAIATIESDKPFLNGSMWGIIGL 120
Db      123 DLRKGVYVPTGKWEGLGDTLVSIHPGPNVTYANAIATIESDKPFLNGSMWGIIGL 182
OY      121 AYAETARPDDSLPEFPDLSLVKQTHVNLFSIQLCGAGFLNLOSEVLASVSGSMIIGIDH 180
Db      183 AYAETARPDDSLPEFPDLSLVKQTHVNLFSIQLCGAGFLNLOSEVLASVSGSMIIGIDH 242
OY      181 SLYTGSLWTPTRRREYVIVIVREINQDLKMKCKEYNYKSLVDSGTTNLRPKRYF 240
Db      243 SLYTGSLWTPTRRREYVIVIVREINQDLKMKCKEYNYKSLVDSGTTNLRPKRYF 302
OY      241 EAAVKSIRKAASSTKRPDGFGLGEOIYVWQAGTTPWNIPEYISLYLMGEVTNOSFRITIL 300
Db      303 EAAVKSIRKAASSTKRPDGFGLGEOIYVWQAGTTPWNIPEYISLYLMGEVTNOSFRITIL 362
OY      301 POQYLRPVEDVATSDDDCYKFAISOSSTGTVMGAVIMGEFYVVDRAKRRIGFAVSACHV 360
Db      363 POQYLRPVEDVATSDDDCYKFAISOSSTGTVMGAVIMGEFYVVDRAKRRIGFAVSACHV 422
OY      361 H 361
Db      423 H 423

```

```

RESULT 14
AAE02595
ID      AAE02595 standard; Protein; 459 AA.
XX
AC      AAE02595;
XX
DT      10-AUG-2001 (first entry)
XX
DE      Human-Asp-2(a) deltatm (His)6 protein.
XX
KW      Human; alpha-secretase; amyloid precursor protein; App; therapy;
KM      Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp2a;
KM      beta-secretase; Asp-2a delta TM; histidine tag; mutant; mutein.
XX
OS      Homo sapiens.

```





|||||  
Db 123 DLKGVVPTQGWEGELGTDLVSIPHGNVTVRANIAITTESDKFFINGSNMEGILGL 182  
QY 121 AYAETARPDDSLPEPFDLSVKOTHPNLFSLQCGAGFPLNQEVLASVGGSMIIGIDH 180  
Db 183 AYAETARPDDSLPEPFDLSVKOTHPNLFSLQCGAGFPLNQEVLASVGGSMIIGIDH 242  
QY 181 SLYTGSLWYTPIRREMYEVIIVRVEINGODLKMDCKEYNTDKSIVDSGTNLRLPKKVF 240  
Db 243 SLYTGSLWYTPIRREMYEVIIVRVEINGODLKMDCKEYNTDKSIVDSGTNLRLPKKVF 302  
QY 241 EAAVKSIRAASTKEFPDGFMLGEOLVCMQAGTTPWNIFPVISLYLMGEVTNOSFRITIL 300  
Db 303 EAAVKSIRAASTKEFPDGFMLGEOLVCMQAGTTPWNIFPVISLYLMGEVTNOSFRITIL 362  
QY 301 POQYLRPEVDVATSDDDCYKFAISQSGTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 360  
Db 363 POQYLRPEVDVATSDDDCYKFAISQSGTGTVMGAVIMEGFYVFDRAKRRIQFAVSACHV 422  
QY 361 H 361  
Db 423 H 423

Search completed: October 30, 2002, 12:27:05  
Job time : 145.577 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 69.1843 Seconds  
(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-75

Perfect score: 1907  
Sequence: 1 MYDNLRRKSGGGYVEMTVG.....VVEDRRKRRIIGFAVSACHVH 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.71.\*  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1907	100.0	501	2 A59090	aspartic proteinase
2	308	16.2	384	2 JC7574	pepsinogen A - Afr
3	308	16.2	385	2 JC7575	pepsinogen A - bul
4	306	16.0	387	2 B38302	pepsin (EC 3.4.23.
5	303	15.9	383	2 JC7573	pepsinogen C - Afr
6	302	15.8	388	1 S19682	pepsin A (EC 3.4.2
7	301	15.8	382	1 PECH	pepsin A (EC 3.4.2
8	298.5	15.7	383	2 A14443	pepsin (EC 3.4.23.
9	298.5	15.7	396	2 A34401	pepsin E (EC 3.
10	296	15.5	387	2 C38302	pepsin (EC 3.4.23.
11	296	15.5	391	2 A43356	pepsin E (EC 3.
12	295.5	15.5	384	2 A39314	pepsin E (EC 3.
13	295.5	15.5	412	1 KHRUD	pepsin D (EC 3.4
14	295	15.5	387	2 D38302	pepsin (EC 3.4.23.
15	287	15.0	407	1 KHRUD	pepsin D (EC 3.
16	286	15.0	387	2 E38302	pepsin D (EC 3.
17	285.5	15.0	387	2 S66465	pepsin (EC 3.4.23.
18	285.5	15.0	444	2 T24204	pepsin E (EC 3.
19	285	14.9	398	1 S19684	pepsin A (EC 3.4.2
20	285	14.9	398	1 S15185	pepsin D (EC 3.
21	283	14.8	388	1 PEMQAR	pepsin A (EC 3.4.2
22	281.5	14.8	381	1 CMSHB	pepsin A (EC 3.4.2
23	281	14.7	388	1 PEMQAR	pepsin A (EC 3.4.2
24	279.5	14.7	410	1 KHRUD	pepsin D (EC 3.
25	278	14.6	386	1 PECH	pepsin A (EC 3.4.2
26	278	14.6	387	2 JC7245	pepsinogen A - com
27	278	14.6	388	1 PEHU	pepsin A (EC 3.4.2
28	277.5	14.6	380	1 A17176	pepsin E (EC 3.4.2
29	277.5	14.6	396	2 S36865	pepsin E (EC 3.

30	277	14.5	388	2 A30142	pepsin A (EC 3.4.2
31	275	14.4	388	2 B30142	pepsin A (EC 3.4.2
32	273.5	14.3	381	1 CMB0	chymosin (EC 3.4.2
33	272	14.3	389	2 JE0371	pepsin C (EC 3.4.2
34	270.5	14.2	377	1 PEMQCT	pepsinogen C (EC 3.4
35	270.5	14.2	389	2 A38302	pepsin (EC 3.4.23.
36	270	14.2	376	2 I45856	aspartic proteinase
37	268.5	14.1	344	1 KHRGD	pepsin D (EC 3.
38	267.5	14.0	381	2 JC7247	pepsinogen C - com
39	266	13.9	405	2 A25379	pepsinogen C (EC
40	265.5	13.9	380	2 S03433	pepsinogen C (EC
41	264	13.8	396	2 T47207	pepsinogen C - com
42	263.5	13.8	388	2 JC7246	pepsinogen C (EC 3.4
43	262.5	13.8	394	2 B43356	pepsinogen C (EC 3.4
44	261.5	13.7	387	2 A45117	pepsinogen C (EC 3.4
45	261.5	13.7	388	2 A29937	pepsinogen C (EC 3.4

ALIGNMENTS

RESULT 1

A59090 aspartic proteinase (EC 3.4.23.-) BACE precursor - human

N:Alternate names: beta-secretase; beta-site APP cleaving enzyme

C:Species: Homo sapiens (man)

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-May-2000

C:Accession: A59090

R:Vaasat, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplio

M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro

Science 286, 735-741, 1999

A:Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran

A:Reference number: A59090; MIMD:20002972

A:Note: submitted to GenBank, September 1999

A:Accession: A59090

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-501 <VAS>

A:Cross-references: GB:AF190725; NID:96118538; PIDN:AF04142.1; PID:96118539

C:Genetics:

A:Gene: BACE

C:Superfamily: beta-secretase

C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-45/Domain: propeptide #status predicted <PRO>

F:46-501/Product: acid proteinase BACE #status predicted <MAT>

F:461-477/Domain: transmembrane #status predicted <TRN>

F:92,288/Active site: Asp #status predicted

F:153,172,223,354/Binding site: carbohydrate (asn) (covalent) #status predicted

F:330-380/Disulfide bonds: #status predicted

Query Match 100.0% Score 1907; DB 2; Length 501;  
Best Local Similarity 100.0%; Pred. No. 5, 5e-154;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MYDNLRRKSGGGYVEMTVGSPQDTNIIVDTGSSNFAVGAAPFLHRYQRLSTYR	60
DB	63	MYDNLRRKSGGGYVEMTVGSPQDTNIIVDTGSSNFAVGAAPFLHRYQRLSTYR	122
QY	61	DLRGGVVPYVTVGGWEGELGDLVSIHGPVNVYRANIAATFESDKFFINGSNWEGITGL	120
DB	123	DLRGGVVPYVTVGGWEGELGDLVSIHGPVNVYRANIAATFESDKFFINGSNWEGITGL	182
QY	121	AYAEIARPDLSLEPFPSLVKQTHVPLVFSLOLCGAGFPLNOSGVLASVGSMTIGIDH	180
DB	183	AYAEIARPDLSLEPFPSLVKQTHVPLVFSLOLCGAGFPLNOSGVLASVGSMTIGIDH	242
QY	181	SLYTGSLWYPIRREMYEIVIVVEINGODLMDCKEYVNDKSIYDSGTTNRLPKVF	240
DB	243	SLYTGSLWYPIRREMYEIVIVVEINGODLMDCKEYVNDKSIYDSGTTNRLPKVF	302
QY	241	EAAVKSTKASSTREKPFDFGWLGEOLVCMQAGTTPNPIFYISLYLMGEVYTNOSFRITIL	300

Db 303 EAAVSKIAKASTEKFPDGFWLGEOVLVCMQAGTTPNNIFPVLSILVMEVTNOSFRITLL 362  
QY 301 PQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGYVVFDRARRKIGFAVACHV 360  
|||  
Db 363 PQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIMEGYVVFDRARRKIGFAVACHV 422  
QY 361 H 361  
|  
Db 423 H 423

## RESULT 2

JC7574  
pepsinogen A - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7574; PC7119  
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.  
J. Biochem. 129, 147-153, 2001  
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens  
A:Reference number: JC7573; MUID:21064922; PMID:11134969  
A:Contents: Stomach  
A:Accession: JC7574  
A:Molecule type: mRNA  
A:Residues: 1-384 <IKU>  
A:Cross-references: DDBJ:AB045380  
A:Accession: PC7119  
A:Molecule type: protein  
A:Residues: 16-35;57-76 <IK2>  
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like  
C:Genetics:  
A:Gene: Pga  
A:Superfamily: pepsin  
C:Keywords: stomach; zymogen

Query Match 16.2%; Score 308; DB 2; Length 384;  
Best Local Similarity 27.5%; Pred. No. 2.6e-18;  
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;

QY 13 YVEMTVGSPPTQTLNIVDTGSSNFVAGAPHPFL-----HRYRQLSSTYRDLRKG 65  
|||  
Db 72 YGTTISIGTPQSTFYVIFDTGSANLWV---PSVYSSQACSHNHNFNPOQSTFGATNTP 128  
QY 66 VVPTQGWEGELGTDVLSIPHGPVTVRANIAITESDK-FFIGSNWEGILGLAYAEI 124  
|||  
Db 129 VSIQGTGSMGFLGVDYQV---GNIQSNQMGFLSESEPSFLYSPFDILGLAFPS 185  
QY 125 IARPDDSLPEPFDVLKQTHVP-NLFSLOLCAGFPLNQSSEVLASVGSMTIIGIDHSLY 183  
|||  
Db 186 IA--SSQATPVDFDMMWSQGLIPQNLFSVLSSDG-----QTGSYVLFGVDNSY 233  
QY 184 TGSIMWTPIRRMVYEVIIIVRVEINGODL-KMDCKEYNDKSIYDSGTNLRLPKVPE 241  
|||  
Db 234 SGLNWPPLAEIYQITLIDSVISINQVACSSQSC-----QAVDTGSLMTGPETPI- 286  
QY 242 AAVKSIKAASTEEKFPDGFWLGEOVLV-CWQAGTTPNNIFPVLSILVMEVTNOSFRITLL 300  
|||  
Db 287 ANIQNTIGASQDSN-----GQYVINCNNISNMPPIV-----TIN 321  
QY 301 PQOYLRPVEDVATSDDCYK-FAISQSSTGT---VMGAVIMEGYVVFDRARRKIGFA 354  
|||  
Db 322 GVOYPLSPSAVYVNOGQSSGFGQAMNLPNGLMILGVDITRQYFYVFDNRANNYATA 381

## RESULT 3

JC7575  
pepsinogen A - bullfrog  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 03-Aug-2001  
C:Accession: JC7575  
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.  
J. Biochem. 129, 147-153, 2001  
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens

A:Reference number: JC7573; MUID:21064922; PMID:11134969  
A:Contents: Stomach  
A:Accession: JC7575  
A:Molecule type: mRNA  
A:Residues: 1-385 <IKU>  
A:Cross-references: DDBJ:AB045376  
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like  
C:Genetics:  
A:Gene: Pga  
A:Superfamily: pepsin  
C:Keywords: stomach; zymogen

Query Match 16.2%; Score 308; DB 2; Length 385;  
Best Local Similarity 27.8%; Pred. No. 2.6e-18;  
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;

QY 13 YVEMTVGSPPTQTLNIVDTGSSNFVAG---AAPHPFLHRYRQLSSTYRDLRKGYYV 68  
|||  
Db 73 YGTTISIGTPQSTFYVIFDTGSSNLMWVPSVYSSQACSHNHNFNPOQSTFGATNTPVSI 132  
QY 69 PYTQGWEGELGTDVLSIPHGPVTVRANIAITESDK-FFIGSNWEGILGLAYAEIAR 127  
|||  
Db 133 QYGTGSMGFLGVDYQV---GNIQNTQIFGLSQSPGSEFLYSPFDILGLAFPSLA- 188  
QY 128 PDDSLPEPFDVLKQTHVP-NLFSLOLCAGFPLNQSSEVLASVGSMTIIGIDHSLYTGS 186  
|||  
Db 189 -SSQATPVDFDMMWSQGLIPQDLFSVYLSQSC---QS-----GSFVLFQGVDFSTYGTN 237  
QY 187 LWPPIRREWYEVIIIVRVEINGODLKM--DCKEYNDKSIYDSGTNLRLPKVFEAAV 244  
|||  
Db 238 LNMVPLAEIYQITLIDSVISINQVACSSGSC-----SAIVDTGSLMTGPETPI--STPI 287  
QY 245 KSIKAASTEEKFPDGFWLGEOVLVCMQAGTTPNNIFPVLSILVMEVTNOSFRITLPQY 304  
|||  
Db 288 ANIQYIGANDQSNQYV---INCNNISNMPPIV-----TINQY 326  
QY 305 LRPVED-VATSDDC---YKFAISQSSTGT---VMGAVIMEGYVVFDRARRKIGFA 354  
|||  
Db 327 PLPASAVYRQSQOSTSGFQAMNLPSTSGDMLTIGVDITREYVYVFDNRANNYVMA 382

## RESULT 4

B38302  
pepsin (EC 3.4.23.-) II-1 precursor - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence\_revision 20-Sep-1991 #text\_change 23-Feb-1997  
C:Accession: B38302  
R:Kageyama, T.; Tanabe, K.; Koizumi, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu  
A:Reference number: A38302; MUID:91009127  
A:Accession: B38302  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:M59235; GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 16.0%; Score 306; DB 2; Length 387;  
Best Local Similarity 27.1%; Pred. No. 3.9e-18;  
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;

QY 13 YVEMTVGSPPTQTLNIVDTGSSNFVAG---AAPHPFLHRYRQLSSTYRDLRKGYYV 68  
|||  
Db 75 YGTTISIGTPQSTFYVIFDTGSSNLMWVPSVYSSQACSHNHNFNPOQSTFGATSTLSI 134  
QY 69 PYTQGWEGELGTDVLSIPHGPVTVRANIAITESD---KFIGSNWEGILGLAYAEI 125  
|||  
Db 135 YTGTSMTGILGIDYQV---GNIEPTNQIFGLSKTEPITFLV--APPDGILGLAYPSI 189  
QY 126 ARPDDSLPEPFDVLKQTHVP-NLFSLOLCAGFPLNQSSEVLASVGSMTIIGIDHSLYT 184  
|||

```
Db 190 SASDAT--PVFDNMKNEGVSEDLFSVYLSNG-----EKGSMVFGIDSSYYT 237
QY 185 GSLWYPIREMYEYEVIVAEINGODLKM--DCKEYNDKSIDSGTTNLRPKVFEA 242
Db 238 GSLNMPVSVHEGYWQIMDSITTINGETIACADSC-----QAVVDGTSTLACPTSAISK 291
QY 243 AVKSIAASSTEFKPPDGFWLGEOLV-CWQAGTTPMNIFFPVSILYMGENVNQSFRITILP 301
Db 292 IQSYIGASKML-----LGENIISCSAIDSLPDIYF-----TINN 325
QY 302 QOYLRPEED-VATSDDC---YFAISQSSTGT--VMGAVIMGEFVVEDRARRKIRGFAV 355
Db 326 VQPLPPLPSAVIILKEDDCISGFGMDLDTSYGELWILGIVFIQYFTVFDRAANNQVGLAA 385
QY 356 SA 357
Db 386 AA 387

RESULT 5
JC7573
pepsinogen C - African clawed frog
N:Alternate names: progastricsin
C:Species: Xenopus laevis (African clawed frog)
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
A:Accession: JG7573; PC7118
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JG7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7573
A:Molecule type: mRNA
A:Residues: 1-383 <IKU>
A:Cross-references: DDBJ:AB045379
A:Accession: PC7118
A:Molecule type: protein
A:Residues: 17-68 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a
C:Gene: Pgc
A:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 15.9%; Score 303; DB 2; Length 383;
Best Local Similarity 27.4%; Pred. No. 6.9e-18;
Matches 104; Conservative 57; Mismatches 122; Indels 96; Gaps 17;

QY 13 YVEVMTVGSPPQTLNIVDTGSSNFAVGA-----APHPLHRYQROLSTYRDLRK 64
Db 67 YVGEISIGTPPQNFVLEFDTGSSNLMVASTYCQSQACTNHP-----FNPSSSTYSSNQO 122
QY 65 GVVYPTOGKWEGLGELTDVLSIPHGNVTVRANIAATESDFINGSN-----WEGILG 119
Db 123 QFSLQGTGSLGILGYDVTI---QNVAISQGEGLSETP-----GTFVYAQGDGLIG 175
QY 120 LAYALARPDDSLPEFPFSLVKQTHVPMFLSLQLCGAGFPINQSEVLASVGSMTIGID 179
Db 176 LAYPSIA--VGGATTVMQMMQO---NLLNQPI--FGFYLSGQS--SQNGGEVAFGVD 225
QY 180 HSLYGSLSWYPIRREMYEYEVIVAEINGOD--LKMCKEYNDKSIDSGTTNLRPK 236
Db 226 QNYITGOIYMTVSETYEWQIGQFSINGQATGWCQSGC-----QAVVDGTSTLAP 279
QY 237 KVFPAAVKSIKASSTEFKPPDGFWLGEOLVCMQAGTTPMNI--FPVYSLYMG----- 288
Db 280 QSVFSSLSIQSICAOQDON-----GQYVYVSCS-----NIQNLPITSLFISGSEFLP 325
QY 289 ---EVTNQS-----FRTITLPOYLRPEVEDVATSDDCIKFAISQSSTGTWGAIVME 338
Db 326 PSAYVLOOSSGCTGTGIMPTLPSQNGQL-----WILGDVFLR 364
QY 339 GFYVVFDRARRKIRGFAVSA 357
```

```
Db 365 EYISYVDLGNNOVGFAATA 383

RESULT 6
S19682
pepsin A (EC 3.4.23.1) 4 precursor - Japanese macaque
N:Alternate names: pepsinogen A isozyme 4
C:Species: Macaca fuscata (Japanese macaque)
C>Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
A:Accession: S19682; S16065
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
Eur. J. Biochem. 202, 205-215, 1991
A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and st
A:Reference number: S19681; MUID:92037645
A:Accession: S19682
A:Molecule type: mRNA
A:Residues: 1-388 <KAG>
A:Cross-references: EMBL:X59753; NID:938070; PIRN:CA042425.1; PID:938071
A>Note: parts of sequence, including amino ends of pepsinogen and activation intermed
C:Comment: This is a minor component of pepsin at all post-partum stages.
C:Superfamily: Although two-step activation is observed, activation is predominantly a o
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di
F:16-388/Product: pepsinogen A 4 #status experimental <PPT>
F:16-62/Domain: activation peptide #status experimental <APT>
F:63-388/Product: pepsin A 4 #status experimental <ENZ>
F:38-39/Cleavage site: Leu-Lys (pepsin) #status experimental
F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental
F:94,277/Active site: Asp #status predicted
F:107-112,268-272,311-344/Disulfide bonds: #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 15.8%; Score 302; DB 1; Length 388;
Best Local Similarity 28.6%; Pred. No. 8.5e-18;
Matches 103; Conservative 58; Mismatches 131; Indels 68; Gaps 15;

QY 13 YVEVMTVGSPPQTLNIVDTGSSNFAVGAHHPFL-----HRYQROLSTYRDLRK 65
Db 76 YVGTIGITPAQNTFTVDFDTGSSNLMV---PSVYCYSLACMDHMLFNPQDSSTYRATSKT 132
QY 66 VVYPTOGKWEGLGELTDVLSIPHGNVTVRANIAATESDFINGSNMGCIIGLAVAE 124
Db 133 VSIYTGTSNMGILGIDYVKY---GGISDTNQIFGLSTEGFLYRAPFGILGLAIPS 189
QY 125 IARPDDSLPEFPFSLVKQTHV-PMLFSLQLCGAGFPINQSEVLASVGSMTIGIDISLY 183
Db 190 IS--SSGATPVFDNIMWQRLVSQDLFSVYLSAD---DQS-----GSVIFGIDSSY 237
QY 184 TGSLSWYPIRREMYEYEVIVAEINGODL--KMKCKEYNDKSIDSGTTNLRPKVFE 241
Db 238 TGSLSWYPIRREMYEYEVIVAEINGODL--KMKCKEYNDKSIDSGTTNLRPKVFE 241
QY 242 AAVKSIAASSTEFKPPDGFWLGEOLV-CWQAGTTPMNIFFPVSILYMGENVNQSFRITIL 300
Db 292 NIQSDIGASENSD-----GEMVYVCSAISLPDIYF-----TIN 325
QY 301 PQYLRPEVEDVATSDDCIK--FAISQSSTGTWGAIVMEGFVVEDRARRKIRGFA 354
Db 326 GVQYPLPPLPSAVIILQSSQCTSGFGQMDVPTESGELWILGIVFIQYFTVFDRAANNQVGLA 385

RESULT 7
PECH
pepsin A (EC 3.4.23.1) precursor - chicken
N:Alternate names: pepsinogen A
C:Species: Gallus gallus (chicken)
C>Date: 18-Apr-1984 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000
A:Accession: JE0370; A00984
R:Sakamoto, N.; Saiga, H.; Yasugli, S.
Biochem. Biophys. Res. Commun. 250, 420-424, 1998
A:Title: Analysis of temporal expression pattern and cis-regulatory sequences of chic
```



```
QY      13 YVENTVGSPPQTLLNIVDTGSSNFVAAGAHPHF-----LHRYQRQLSSTYYRLDKRG   65  
       |::|::|||::||||| | | : ||::|::|:  
Db     75 YFGTISIGTFPQDFVIYFDTGSSNLNW---PSITCSSLACALHKRPNEDESSSYQGTSYT 133
```

QY	13	YIYEMVYGSPPOTLNIYVDGSSNFANGA----	APHFLLHRYQROLSSTYDLKRGYV	68
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	74	YFETISIGSPONFVIEFDGSSNLMPVSUYCTSPACQTHPRVPHSLSTSYREVNCSFSI		133
QY	69	PIYQGWEGELGDLVSI:PHGNPTVARNALAIRESKFFINSNMEGLLGLAYELIAR	122	
Db	134	QYGTGSLTGLIADGVSV-EGLTIVGQGFGEBSVEPKRTVH-AEFDGLLIGVPSLA-	190	
QY	129	DSLEIPEFSLVKQTHVNFLESLDLCAGFPLNOSVLIASVGSMTIGIDHSLTGSLL	188	
Db	191	-GGVTVFEDNMAQ-----NLVALPR-----FSYVMSSNPGSGSGELTFGGIYDPSHFGSLN	242	
QY	189	YTPIREMYEVIIIVVEINQDLMKQCEKYNDRKSIVDSGTLNLRPKRYEAAVKSIR	248	
Db	242	WVIVTQKAWYALDGLIGVQ--DSVMCESE--CGQALVDIGTSLIGP-----PGKIRLQ	293	
QY	249	AASSTKEPFDGWLBEOLVCMAQAGTTPMNIPEVLSILYMGENTNOSFRI----	307	
Db	294	EALGATVYDEG-----SVQC-----ANINMMIDVT--FIINGVPTLNPTA	333	
QY	304	YLRPEVDVATQDDDCYKRAISQSTG-----TYMGAVIMEFYVFPBARKR	350	
Db	334	Y--TLIDFVDGMQVC-----STGEGGLEIOPRPAAGPLMLIGDVFIRQFAVPEGRGNR	383	
QY	351	IGFA	354	

Db 384 VGLA 387

## RESULT 12

A39314

gastricsin (EC 3.4.23.3) precursor - bullfrog

C:Species: Rana catesbeiana (bullfrog)

C:Date: 19-Jun-1992 #sequence,revision 19-Jun-1992 #text,change 22-Jun-1999

C:Accession: A39314

R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageya

J. Biol. Chem. 266, 22436-22443, 1991

A:Title: Purification, characterization, and amino acid sequences of pepsinogens and pep

A:Reference number: A39314; MUID:92042186

A:Accession: A39314

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-384 &lt;YAK&gt;

A:Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688

A:Superfamily: pepsin

C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 15.5%; Score 295.5; DB 2; Length 384;

Best Local Similarity 26.1%; Pred. No. 3e-17;

Matches 98; Conservative 59; Mismatches 130; Indels 89; Gaps 15;

13 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

14 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

15 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

16 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

17 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

18 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

19 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

20 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

21 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

22 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

23 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

24 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

25 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

26 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

27 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

28 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

29 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

30 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

31 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

32 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

33 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

34 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

35 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

36 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

37 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

38 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

39 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

40 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

41 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

42 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

43 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

44 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

45 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

46 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

47 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

48 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

49 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

50 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

51 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

52 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

53 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

54 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 65

A:Reference number: S30749; MUID:87231068

A:Accession: S30749

A:Molecule type: mRNA

A:Residues: 1-412 &lt;MES&gt;

A:Cross-references: EMBL:X05344; NID:g29677; PIDN:CAA28955.1; PID:g29678

R:May, F.E.B.; Smith, D.J.; Westley, B.R.

Gene 134, 277-282, 1993

A:Title: The human cathepsin D-encoding gene is transcribed from an estrogen-regulate

A:Reference number: PC2066; MUID:94085791

A:Accession: PC2066

A:Molecule type: DNA

A:Residues: 1-23 &lt;MAX&gt;

A:Cross-references: GB:L12980; NID:g291930; PIDN:AAA16314.1; PID:g455429

A:Experimental source: MCF-7 cell

R:Cavallies, V.; Augereau, P.; Rochefort, H.

Proc. Natl. Acad. Sci. U.S.A. 90, 203-207, 1993

A:Title: Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate

A:Reference number: 159236; MUID:93126342

A:Accession: 159236

A:Status: translation not shown; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-22 &lt;CAV1&gt;

A:Cross-references: GB:S5257; NID:g263124; PIDN:AAID13868.1; PID:g4261568

R:Augereau, P.; Miralles, F.; Cavallies, V.; Gaudelet, C.; Parker, M.; Rochefort, H.

Mol. Endocrinol. 8, 693-703, 1994

A:Title: Characterization of the proximal estrogen-responsive element of human cathep

A:Reference number: 157716; MUID:95021301

A:Accession: 157716

A:Status: translation not shown; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-22 &lt;CAV2&gt;

A:Cross-references: GB:S74689; NID:g786350; PIDN:AAID14156.1; PID:g4261856

R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.

submitted to the Brookhaven Protein Data Bank, April 1993

A:Reference number: A51839; PDB:1LVA

A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 65-161,170-241

R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.

submitted to the Brookhaven Protein Data Bank, April 1993

A:Reference number: A51840; PDB:1LVB

A:Contents: annotation; X-ray crystallography, 2.5 angstroms, with inhibitor residues

R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Hosur, M.V.; Sowder II, R.C.; Cachau, R.E.;

Proc. Natl. Acad. Sci. U.S.A. 90, 6796-6800, 1993

A:Title: Crystal structures of native and inhibited forms of human cathepsin D: Impli

A:Reference number: A48229; MUID:93342076

A:Contents: annotation; X-ray crystallography, 2.5 angstroms

C:Comment: cathepsin D is a ubiquitous lysosomal proteinase.

C:Comment: In addition to the propetide, residues 163-168 and 411-412 are proteolyti

C:Comment: The carbohydrate bound to 134-asn contains a mannose-6-phosphate that is b

C:Genetics: GDB:CTSD

A:Gene: GDB:CTSD

A:Cross-references: GDB:120512; OMIM:116840

A:Map position: 11p15.5-11p15.5

C:Function:

A:Description: limited specificity endopeptidase

A:Pathway: intracellular protein degradation

C:Superfamily: pepsin

C:Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradati

F:1-20/Domain: signal sequence #status predicted &lt;SIG&gt;

F:21-64/Domain: propeptide #status predicted &lt;PRO&gt;

F:65-162/Domain: propeptide #status predicted &lt;PRO&gt;

F:163-168/Domain: propeptide #status predicted &lt;PRO&gt;

F:169-410/Region: phosphotransferase recognition

F:411-412/Region: phosphotransferase recognition

F:91-160,110-117,286-290,329-366/Duplicate bonds: #status experimental

F:97,295/Active site: Asp #status experimental

F:134,263/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 15.5%; Score 295.5; DB 1; Length 412;

Best Local Similarity 28.5%; Pred. No. 3e-17;

Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;

13 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

14 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

15 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

16 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

17 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

18 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

19 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66

20 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHPFL-----HRYQRLSSTYRDLKRG 66



```

0Y 67 YVPYQGMKEGSLGTDIVSI-----HGPNVTARNIAITRESDFLINGSNMEGI 117
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 DIHYSGSLISGLSDIYVSPCCQSSASALGCGVVERKVEGEATKQPCITFIIAKFPDI 198
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 118 IGLAYAEIARPPDLSLEPFDSLVKQTHV-PNLFSLQCGAGFPLNQEVLASVGSMTIG 176
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 IGMAYPRIS--VNNVLPEVDNIMQCKLVDQNFISFYL-----SRPDQAQGGELMIG 248
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 177 GIDHSLYTGSLMYTPIIRRMWYEVILIVREI-NGODLKMDCKEYNDSYDNGTINRL 235
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 GIDTSYTYKGSLSYLVNTRKATWQVHLDQVEVASGLTL--CE--GCEIYDVTGSLMWG 303
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 236 PKKVEAAVKSITKASSTEKEFPDGFMLGEQLV-CWQAGTTPWNIPIVLSYLMGEVNTS 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 PVDVEVRELQKAGIAPLAIQ-----GEYMIPECKEYR-----LPATILKLG---KG 346
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 295 FRITILLPOOYLRPVDDVATSDODCYKFAISQ-----SSGTVMGAVIMEGFVYVDRARK 349
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 YKLS--PEDIYTLAKVQAQKTL--CLSGFMGMDIPPESGFMTLGDVFIGRITYVDRONN 402
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 350 RIGFAVSA 357
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 RVGFAEAA 410
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 14
D38302
pepsin (EC 3.4.23.-) II-4 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: D38302
R:Kageyama, T.; Tanabe, K.; Kolwail, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucle
A:Reference number: A38302; MUID:91009127
A:Accession: D38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic protease; hydrolase; phosphoprotein; protein digestion

Query Match          15.5%, Score 295, DB 2; Length 387;
Best Local Similarity 26.1%, Pred. No. 3.3e-17;
Matches 97; Conservative 65; Mismatches 123; Indels 86; Gaps 14;

```

```

0Y 13 YVVEVNTVSSPOTLNIILVDTGSSNFVAGAAHPHF-----LHRVYQJOLSTYIDLKRG 65
Db 11 : :::::|||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
75 YEGTISITPPODFTVVIDTGGSNLWV--PSTYCSSLACALHKRFNEDSSTYOGTSET 131
QY 66 VVVPYTOGRMEGELCTDVLSPHCPNVTVRANIAITAITESDKFF-----INGSWME 115
Db 132 LSTVGTGSMNGILGIDRV-----KVGSIEDTNQIFGLSKTEPGLTFLFAFD 179
QY 116 GILGLAVAEIARPDSDLEPFDFSLVKQTHV-PNLFSLDLCAGPFLNOSVLA5VGSMT 174
Db 180 GILGLAVPSSISDDT--PVEDNMNMNEGLVSODLF5VYLLSDD-----EKSGSLVM 227
QY 175 IGGIDHSLYTGSMLWTPPIRREMYEVLIVRVEINQODLKM--DCKEYNYDKSIVDSGTTN 232
Db 228 FGGIDSSYVTSGLMNVPRSYEGYMQITMDSV5INGETIACDSC-----QAIYDTG5SL 281
QY 233 LRLPKKVEAAVKSITKAASSTEEKFPDGFWLGEOLV-CWQAGTTPWNIPFV5ISLYLMGEVT 291
Db 282 LTGP--TSAISNIQSYIGASK-----NLLEBNV5ISCAISDLP5IV----- 321
QY 282 NOSFRITLIPQOYLPEVEDVAT5ODDCKFAISQ5ST-----VMGAVIMEGRVYVD 345
Db 322 -----TINGIQYPLPA5AVILKEDDCTSGLEGHNVDTYIGELMILD5IVFROI5FVD 375
QY 346 RARRKRGFAVS 356
11 : :: : : :

```

Db 376 RANNQGLAA 386

## RESULT 15

cathepsin D (3.3.4.23.5) precursor - rat  
C:Species: *Rattus norvegicus* (Norway rat)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Jun-1999  
C:Accession: S1311; C31918; J01177; PQ0222  
C:Birth, N.P.; Loh, Y.P.  
Nucleic Acids Res. 18, 6445-6446, 1990  
A:title: Cloning, sequence and expression of rat cathepsin D.  
A:Reference number: S1311; MUID:91057150  
A:Accession: S1311  
A:Molecule type: mRNA  
A:Residues: 1-407 <BIR>  
A:Cross-references: EMBL:X54467; NID:955881; PIDN:CAA38349.1; PID:955882  
R:Yonezawa, S.; Takahashi, T.; Wang, X.; Wong, R.N.S.; Hartsuck, J.A.; Tang, J.  
J. Biol. Chem. 263, 16504-16511, 1988  
A:title: Structures at the proteolytic processing region of cathepsin D.  
A:Reference number: A92681; MUID:89034127  
A:Accession: C31918  
A:Molecule type: protein  
A:Residues: 154-162, 'T', 164-170 <YON>  
R:Fujita, H.; Tanaka, Y.; Noguchi, Y.; Kono, A.; Himeno, M.; Kato, K.  
Biochem. Biophys. Res. Commun. 179, 190-196, 1991  
A:title: Isolation and sequencing of a cDNA clone encoding rat liver lysosomal cathep  
A:Reference number: J01177; MUID:91354249

A:Molecule type: mRNA  
A:Residues: 1-14, 'A', 16-204, 'N', 206-261, 'N', 263-407 <EU>  
A:Accession: P00222  
A:Molecule type: Protein  
A:Residues: 65-74;118-127;165-174 <FU>  
A:Experimental source: liver  
C:Comment: Cathepsin D in rat liver lysosome occurs as a mixture of both a single chain  
C:Function:  
A:Description: limited specificity endopeptidase  
A:Pathway: intracellular protein degradation  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradation  
F:1-20/Domain: signal sequence #status predicted <IG>  
F:21-64/Domain: propeptide #status predicted <PX>  
F:65-407/Product: cathepsin D, 43k single-chain form #status predicted <MA>  
F:65-164/Product: (or 65-165) cathepsin D 12k light chain #status predicted <MA2>  
F:65-117/Product: (or 65-118) cathepsin D 9k light chain #status predicted <MA4>  
F:118-407/Product: cathepsin D 34k heavy chain #status predicted <MA5>  
F:165-407/Product: (or 166-407) cathepsin D 30k heavy chain #status predicted <MA3>  
F:91-160,110-117,281-285,324-361/Disulfide bonds: #status predicted  
F:97,290/Active site: Asp #status predicted  
F:134,258/Binding site: carboxylate (Asp) (covalent) #status predicted

Query Match	15.0%;	Score 287;	DB 1;	Length 407;
Best Local Similarity	27.28;	Pred. No. 1,7e-16;		
Matches 101;	Conservative 64;	Mismatches 135;	Indels 72;	Gaps 16;
QY	13	YVENVTVGSPQTLNIILVDGSSNFAYGAAPHPL-----HRYQROLSTYRDLKRGV	66	
		:::		
Db	79	YVGEIGISTPPOCFVVDVETDSSNLWAPSIICKLLDLACAWHHRKYNDSKSTYVKNCTSF	138	
QY	67	YVPYTGQWBEELGTDLVSIPIHGPNVYVRAIMAIITESDKF-----INGSNWEG	116	
		:::		:::
Db	139	DIHGGSGSLSTLSODIVSVP-----CKSDLGIGIKVEKQIFEGATKPGGVYFAAFEDG	192	
QY	117	ILGLAYAEIARPDLSLEFFEDSLVKQTHV--PNEISLQDLCGAGPFLNOSVLAYSGSMII	175	
		:::		:::
Db	193	ILGKGPYFIS--VANKVLPVFEDNLKKOKLVERNIPS-----FYLNK-DPTGQGGGGLML	242	
QY	176	GGISHLSLTGSLMTPIPRREMYEVIIVRVINODLAKMCKEYINRKSYIDSTTTLRL	235	
		:::		:::
Db	243	GGTISRYYHGELSTLVNTKRAIYQWVHMDQLEV--GSELTLL-CK--GGCEAIVDTSTSLVG	298	
QY	236	PKKVEAAVKSIRKAASTTEKPPDGFWMLEQDLY--CWAQATPPWNIFFPISILYLMGEVYNOS	294	

```
Db      299  PVDEYKEIQAIGAVPLIQ-----GEYMIPEKVS-----LPITFKLGQ----- 340
QY      295  FRITILPOOYLRPYEDVATSODDCYKFAIS-----QSTGTVMGAVIMEGFYVED 345
Db      341  -NYELHPEKYLILKYSQAGKT-----ICLSGFMGMDIPPSGPLWILGDFVFIGCYTVVED 393
QY      346  RAKKRIGFVSA 357
Db      394  REYNREVGFAKAA 405
```

Search completed: October 30, 2002, 12:31:28  
Job time : 70.1843 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:20:50 ; Search time 3.20393 Seconds  
(without alignments)  
277.344 Million cell updates/sec

Title: US-09-724-571-81

Perfect score: 34

Sequence: 1 EVMXVAF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	94.1	8	21	AA07872
2	27	79.4	7	21	AA07871
3	27	79.4	347	22	AA03166
4	26	76.5	107	20	AA135564
5	26	76.5	216	21	AA195715
6	26	76.5	218	22	AA063744
7	26	76.5	254	20	AA135214
8	26	76.5	271	20	AA152227
9	26	76.5	271	21	AA082005
10	26	76.5	271	22	AA064357
11	26	76.5	271	22	AA092851

12	26	76.5	271	22	AA088446
13	26	76.5	276	21	AA042611
14	26	76.5	282	21	AA057116
15	26	76.5	559	17	AA006619
16	26	76.5	579	22	AA096355
17	26	76.5	587	21	AA074285
18	26	76.5	587	21	AA074287
19	26	76.5	587	21	AA074288
20	26	76.5	821	20	AA034479
21	26	76.5	869	20	AA034354
22	26	76.5	1249	22	AA071313
23	25	73.5	14	21	AA007888
24	25	73.5	27	22	AA028205
25	25	73.5	27	22	AA033380
26	25	73.5	27	22	AA018839
27	25	73.5	27	22	AA054165
28	25	73.5	27	22	AA065559
29	25	73.5	27	22	AA014432
30	25	73.5	27	22	AA026845
31	25	73.5	27	22	AA002159
32	25	73.5	54	19	AA079151
33	25	73.5	54	20	AA081408
34	25	73.5	88	22	AA056968
35	25	73.5	135	22	AA026145
36	25	73.5	144	22	AA031774
37	25	73.5	152	22	AA096397
38	25	73.5	189	22	AA052817
39	25	73.5	225	22	AA065926
40	25	73.5	237	22	AA090704
41	25	73.5	280	22	AA090751
42	25	73.5	443	21	AA042806
43	25	73.5	453	21	AA068799
44	25	73.5	479	20	AA093603
45	25	73.5	482	22	AA031776

#### ALIGNMENTS

##### RESULT 1

ID AAB07872 standard; peptide: 8 AA.

AC AAB07872;

XX 14-NOV-2000 (first entry)

XX A beta-secretase inhibitor peptide.

DE Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.

XX Synthetic.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 4 /note="hydroxyethylene"

XX WO200047618-A2.

XX 17-AUG-2000.

XX 10-FEB-2000; 2000WO-US03819.

XX 10-FEB-1999; 99US-0119571.

XX 15-JUN-1999; 99US-0139172.

XX (ELAN-) ELAN PHARM INC.

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
FI Slnha S, Tatsuno G, Tung J, Wang S, McConlogue L;

DR WPI: 2000-533011/48.  
 XX  
 PT Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX  
 PS Disclosure; Page 12; 121pp; English.  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents an inhibitor of beta-secretase enzyme.  
 CC  
 SO Sequence 8 AA:  
 Query Match 94.1%; Score 32; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 6,4e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 EVMXVAEF 8  
 1 EVMXVAEF 8  
 DB 1 EVMXVAEF 8  
 RESULT 2  
 AAB07871  
 ID AAB07871 standard; peptide; 7 AA.  
 AC  
 XX AAB07871;  
 AC  
 XX 14-NOV-2000 (first entry)  
 DT  
 XX  
 DE A beta-secretase inhibitor peptide.  
 XX  
 XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 3 /note="hydroxyethylene"  
 FT  
 XX  
 PV WO200047618-A2.  
 XX  
 PD 17-AUG-2000.  
 PD  
 XX  
 PF 10-FEB-2000; 2000WO-US03819.  
 PF  
 XX 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 PR  
 XX  
 PA (ELAN-) ELAN PHARM INC.  
 PA  
 XX  
 PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX  
 DR WPI: 2000-533011/48.  
 XX  
 PT Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX  
 PS Disclosure; Page 12; 121pp; English.  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves

CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents an inhibitor of beta-secretase enzyme.  
 CC  
 SO Sequence 7 AA:  
 Query Match 79.4%; Score 27; DB 21; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 6,4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 VMXVAEF 8  
 1 VMXVAEF 7  
 DB 1 VMXVAEF 7  
 RESULT 3  
 AAG93168  
 ID AAG93168 standard; Protein; 347 AA.  
 AC  
 XX AAG93168;  
 AC  
 XX 26-SEP-2001 (first entry)  
 DT  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 6922.  
 DE  
 XX Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 KW  
 XX Corynebacterium glutamicum.  
 OS  
 XX  
 PN EPI108790-A2.  
 PN  
 XX 20-JUN-2001.  
 PD  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 PR  
 XX  
 PA (KYOW) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 PI  
 XX  
 DR WPI: 2001-376931/40.  
 DR  
 DR N-PSDB; AAH68387.  
 DR  
 XX  
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 PS Claim 17; SEQ ID NO: 6922; 246bp + Sequence Listing; English.  
 PS  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

SQ Sequence 347 AA:

Query Match 79.4%; Score 27; DB 22; Length 347;

Best Local Similarity 62.5%; Pred. No. 1.8e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I : I I I I  
Db 78 EILIVAEF 85RESULT 4  
AAV35564

ID AAV35564 standard; Protein; 107 AA.

XX AAV35564;

AC AAV35564;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae protein not found in C. trachomatis.

KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

XX vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

XX W09927105-A2.

PM 03-JUN-1999.

PD 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97ER-0014673.

XX (GEST ) GENSET.

PA Grifffals R;

PI WPI: 1999-357842/30.

DR Genome sequence of Chlamydia pneumoniae

PT Page 1307; Disclosure; 1912pp; English.

XX AAV34584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

CC Sequence 107 AA;

Query Match 76.5%; Score 26; DB 20; Length 107;

Best Local Similarity 62.5%; Pred. No. 93;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I : I I I I  
Db 91 EVMRIARF 98RESULT 5  
AAV95715

ID AAV95715 standard; Protein; 216 AA.

XX AAV95715;

AC 25-OCT-2000 (first entry)

DT Cosmid cHRIM5 encoded protein P21-7r.

DE Cosmid cHRIM5; nematocide; nematode; biological control agent;

KM transgenic plant; helminthiasis; P21-7r.

XX Xenorhabdus bovienii.

OS W0200042855-A1.

XX 27-JUL-2000.

PD 24-JAN-2000; 2000WO-GB00219.

XX 22-JAN-1999; 99GB-0001499.

PR (HORT-) HORTICULTURE RES INT.

PA Morgan JAW, Jarrett P, Ellis D, Ousley MA;

XX WPI: 2000-499157/44.

DR N-PSDB; AAA50029.

PT Novel composition used to control parasitic nematodes, especially in

XX plants such as maize, cotton, soya, and rice, comprises a bacterium

XX which is a symbiont of an entomopathogenic nematode -

XX Example 6; Page 44; 74pp; English.

XX The present sequence is that of protein P21-7r encoded by an open

CC reading frame identified in cosmid cHRIM5 (see AAA50029). cHRIM5 was

CC obtained by ligating Xenorhabdus bovienii strain I73 (NCIMB 40986)

CC Sau3A-digested DNA fragments into the BamHI site of the Stratagene

CC cosmid vector SupercosI, packaging into Escherichia coli XL Blue I,

CC and screening for nematocidal activity against Caenorhabditis elegans.

CC Analysis of the DNA indicated a number of open reading frames for

CC which the corresponding protein sequences were determined (see

CC AAV95685-Y95735). Nematodes can be controlled through the use of

CC bacteria associated symbiotically with an entomopathogenic nematode.

CC Such bacteria include Xenorhabdus and Photobacterium spp. such as X.

CC bovienii strain I73. The symbiotic bacteria, an engineered

CC bacterium, or a nematocidal protein obtained from such bacteria,

CC can be used to control helminthiasis in a human or domesticated

CC animal or for the control of plant pathogen nematodes. Also

CC claimed are vectors for expressing nematocidal proteins in host

CC cells, and transgenic plants.

CC Sequence 216 AA;

Query Match 76.5%; Score 26; DB 21; Length 216;

Best Local Similarity 62.5%; Pred. No. 1.9e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8  
I : I I I I  
Db 108 EIMGVANF 115RESULT 6  
AAB63744

ID AAB63744 standard; Protein; 218 AA.

XX AAB63744;

AC 26-MAR-2001 (first entry)

DE Human prostate cancer associated antigen protein sequence SEQ ID NO:1106.

KW Human: breast cancer; gastric cancer; prostate cancer; diagnosis;  
 KW cancer associated antigen; cytostatic; cancer vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC200073801-A2.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 26-MAY-2000; 2000WO-US14749.  
 XX  
 PR 28-MAY-1999; 99US-0136526.  
 PR 10-SEP-1999; 99US-0153454.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Obata Y;  
 XX  
 DR WPI: 2001-025274/03.  
 XX  
 PT Nucleic acids encoding breast, gastric and prostate cancer associated  
 PT antigen precursors, useful for diagnosing and treating a condition  
 PT characterized by expression of an abnormal amount of a protein, e.g.  
 PT cancer -  
 XX  
 PS Example 1: Page 696-697; 799pp; English.  
 CC  
 CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014  
 CC represent nucleotide sequences encoding human breast, gastric and  
 CC prostate cancer associated antigen precursors (CAAP) respectively.  
 CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63970  
 CC represent human breast, gastric and prostate CAAP protein sequence  
 CC respectively. CAAPs have cytostatic activity and can be used in the  
 CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic  
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a  
 CC condition characterised by expression of an abnormal amount of a protein,  
 CC e.g. cancer.  
 CC  
 SQ Sequence 218 AA:  
 Query Match 76.5%; Score 26; DB 22; Length 218;  
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 EYMXVAEF 8  
 |||  
 Db 167 EVKNVAEF 174  
 RESULT 7  
 AAY35214  
 ID AAY35214 standard: Protein; 254 AA.  
 AC AAY35214;  
 XX  
 DF 13-SEP-1999 (first entry)  
 XX  
 DE Amino acid sequence of a Chlamydia pneumoniae protein.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.

XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffais R;  
 XX  
 DR WPI: 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Page 1062-1063; Disclosure; 1912pp; English.  
 CC  
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 CC  
 SQ Sequence 254 AA:  
 Query Match 76.5%; Score 26; DB 20; Length 254;  
 Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 EYMXVAEF 8  
 |||  
 Db 204 ELLAIAEF 211  
 RESULT 8  
 AAY15227  
 ID AAY15227 standard: Protein; 271 AA.  
 AC AAY15227;  
 XX  
 DF 26-OCT-1999 (first entry)  
 XX  
 DE Human receptor protein (HURP) 6 amino acid sequence.  
 XX  
 KW receptor; cancer; autoimmune disorder; inflammation;  
 KW antagonist; cell surface protein; cell signalling;  
 KW antibody; human receptor protein; HURP; reproductive disorder;  
 KW developmental disorder; gastrointestinal disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FH Misc-difference 63  
 FT /note= "Potential CAMP-/cGMP-dependent protein-  
 FT kinase phosphorylation site"  
 FT  
 FT Misc-difference 95  
 FT /note= "Potential casein kinase II-  
 FT phosphorylation site"  
 FT  
 FT Misc-difference 114  
 FT /note= "Potential casein kinase II-  
 FT phosphorylation site"  
 FT  
 FT Misc-difference 213  
 FT /note= "Potential casein kinase II-  
 FT phosphorylation site"  
 FT  
 FT Misc-difference 6  
 FT /note= "Potential protein kinase C-  
 FT phosphorylation site"  
 FT  
 FT Misc-difference 25  
 FT /note= "Potential protein kinase C-  
 FT phosphorylation site"  
 FT  
 FT Misc-difference 59  
 FT /note= "Potential protein kinase C-  
 FT phosphorylation site"

FT	Misc-difference 63	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 75	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 123	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 135	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 189	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 203	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 247	/note=	"Potential protein kinase C-phosphorylation site"
FT			
FT	Misc-difference 160	/note=	"Potential tyrosine kinase-phosphorylation site"
FT			
FT	Binding-site 71..78	/note=	"Potential ATP/GTP-binding site-motif A (P-loop)"
FT			
PN	WO9941375-A2.		
XX			
XX	19-AUG-1999.		
XX			
XX	05-FEB-1999;	99WO-US02572.	
XX			
XX	12-FEB-1998;	98US-0022939.	
XX			
PA	(INCYTE) INCYTE PHARM INC.		
XX			
PI	Au-Young JI, Bandman O, Baughn M, Corley NC, Guegler KJ;		
PI	Hillman JL, Lal P, Shah P, Tang YF, Yue H;		
XX			
DR	WPI: 1999-494536/41.		
DR			
XX	N-PsDB: AA06371.		
XX			
PT	New human receptor proteins, used e.g. to treat, prevent and		
PT	diagnose gastrointestinal and developmental disorders - and related		
PT	nucleic acids, vectors, transformed cells, antibodies, agonists and		
PT	antagonists		
XX			
XX			
PS	Claim 1; Page 83; 94pp; English.		
XX			
CC	The Human receptor protein 6 (HURP-6) and mouse signal recognition		
CC	particle beta subunit share 90% identity.		
CC	HURP-6 is expressed in cancerous, inflamed, reproductive and gastro-		
CC	intestinal tissue. HURP-6 therefore appears to have a role in		
CC	cancer, autoimmune/inflammatory disorders, reproductive disorders,		
CC	and gastrointestinal disorders.		
CC	This gives rise to the possibility of using an antagonist or an antibody		
CC	of HURP to treat or prevent cancer or autoimmune/inflammatory disorders.		
XX			
XX	Sequence 271 AA;		

Query Match	76.5%	Score 26;	DB 20;	Length 271;
Best Local Similarity	75.0%;	Pred. No. 2.4e+02;		
Matches	6;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0;
QY	1	EVMAVVEF	8	
Db	151	EYKDYAEF	158	
RESULT 9				
AAB28205				

ID	AA28205 standard; Protein; 271 AA.
XX	
AC	AA28205;
XX	
DT	30-JAN-2001 (first entry)
XX	
DE	Novel human protein #3.
XX	
KW	Cytostatic; vaccine; human; breast tumour; antigen; breast cancer.
XX	
OS	Homo sapiens.
XX	
PN	W0200052165-A2.
XX	
PD	08-SEP-2000.
XX	
PF	29-FEB-2000; 2000MO-US05431.
XX	
PR	04-MAR-1999; 99US-0262505.
PR	19-MAR-1999; 99US-0272886.
PR	17-SEP-1999; 99US-0396313.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Lodes MJ;
XX	
DR	WPI; 2000-572184/53.
DR	N-PSDB; AAC69684.
XX	
PT	Breast tumor antigen polypeptides and polynucleotides, useful for
PT	manufacturing vaccines and compositions for treating, diagnosing, and
PT	monitoring breast cancer -
XX	
PS	Example; Fig 2; 140bp; English.
XX	
CC	The present invention relates to immunogenic portions of new human
CC	breast tumor antigens (AA28183-B28214) and their coding sequences
CC	(AAC69645-C69604). The breast tumor antigen polypeptides of the present
CC	invention and their coding sequences are useful for inhibiting the
CC	development of breast cancer in a patient. The breast tumor antigen
CC	polypeptides and polynucleotides may be used in vaccines and
CC	pharmaceutical compositions for treating breast cancer, and for
CC	diagnosing and monitoring the cancer. The present sequence is a
CC	immunogenic portion for one such human breast cancer tumor antigen.
SQ	Sequence 271 AA;

```

Query Match          76.5%;   Score 26;   DB 21;   Length 271;
Best Local Similarity 75.0%;   Pred. No. 2.4e+02;
Matches      6;   Conservative      0;   Mismatches      2;   Indels      0;   Gaps      0;

Qy      1 EVMXVAEF 8
      || |||
      151 EYKDVAEF 158

Db

RESULT 10
AAG64357
ID      AAG64357 standard; Protein; 271 AA.
XX
XX      AAG64357;
AC
XX
XX      01-OCT-2001 (first entry)
DT
XX
XX      Human signal recognition particle receptorbeta.
DE
XX
XX      Human; signal recognition particle receptor beta; SRPBeta.
KW
XX
XX      Homo sapiens.
OS
XX
XX      CNL279290-A.
FN
XX
XX      10-JAN-2001.
PD

```

XX 23-JUN-1999; 99CN-0108547.  
 XX  
 PR 23-JUN-1999; 99CN-0108547.  
 XX  
 PA (UYFU-) UNIV FUDAN.  
 XX  
 PI Yu L, Fu Q, Zhao Y;  
 XX WPI: 2001-266742/28.  
 DR N-PSDB; AAH73878.  
 XX  
 PT New human signal recognition particle receptor beta nucleic acid for  
 PI preparing the protein encoded by it -  
 XX  
 PS Claim 2; Pages 15-16 (Disclosure); 20pp; Chinese.  
 XX  
 XX The present sequence is the protein sequence for human signal recognition  
 CC particle receptorbeta (SRPbeta). The present protein is the homolog of  
 CC mouse SRP beta. Application of human SRPbeta coding sequence and  
 CC protein, and their preparing process are also disclosed.  
 XX  
 SQ Sequence 271 AA;

Query Match	76.5%	Score 26;	DB 22;	Length 271;
Best Local Similarity	75.0%;	Pred. No.	2.4e+02;	
Matches	6;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0;

QY	1	EVMXVAEF	8
Db	151	EVKDVAEF	158

RESULT	11
AAB92851	
ID	AAB92851 standard; Protein; 271 AA.

AC AAB92851;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:11414.

KW	Human; primer; detection; diagnosis; aurisense therapy; gene expression
XY	

OS Homo sapiens.  
XX

EP10/4017-A2  
PN  
XX

XX  
ED  
S

XX

PR 27-AUG-1999; 99JP-0300253

PR 02-MAY-2000; 2000JP-0183767  
00-TYN-2000; 2000TP-0241899

XX  
DA (HET T-) HET TY RES INST

XX  
PI Ota T, Tsodai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J

PI Ishii S, Sugiyama I, Watanabe H, Nagai M  
XX

WFL; 2001-210/42/24  
DK  
XX

full-length cDNAs defined in the specification, and for the detection

full-length cDNAs -

PS Claim 8; SEQ ID 11414; 2537pp + CD ROM; English

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

Query Match	76.5%	Score 26;	DB 22;	Length 271;
Best Local Similarity	75.0%	Pred. No. 2.4e+02;		
Matches	6;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0

QY	1	EVMXVAEF	8
Db	151	EVKDVAEF	158

RESULT 12  
AAB88446  
ID AAB88446 standard; Protein; 271 AA

AA AAB88446;  
AC

DT 23-MAY-2001 (first entry)

DE Human membrane or secretory protein clone PSEC0230

KW Human; secretory protein; membrane protein; vaccine, gene therapy, rheumatoid arthritis: diabetes

XX  
XX  
Homo sapiens

XX  
PN  
EP1067182-A2.

XX  
PD 10-JAN-2001.

AA  
PF 07-JUL-2000; 2000EP-0114090.

08-JUL-1999; 99JP-0194179.

02-MAY-2000; 2000JP-0183766.  
PR

PA (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa I, Kawada I, Sugiyama Y

DR WPI; 2001-0933989/11.  
 DR N-PsDB: AAF93873.

XX Nucleic acids encoding secretory proteins/membrane proteins, useful in

**Pt**      gelle therapy or as candidate cases for treatment.

P3 CLAIM 1, DEQ ID NO. / OFFICE

XX XX

which encode human secretory or membrane proteins represented by



CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretion  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
 CC (ELISA). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.

XX Sequence 271 AA;

Query Match 76.5%; Score 26; DB 22; Length 271;  
 Best Local Similarity 75.0%; Pred. No. 2.4e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVKXVAEF 8  
 II IIII  
 Db 151 EVKDAEF 158

RESULT 13

AAB42611  
 ID AAB42611 standard; Protein; 276 AA.

XX AAB42611;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2375 polypeptide sequence SEQ ID NO:4750.

Human: Open reading frame; ORFX; detection; cytosolic; hepatotropic;  
 human; vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
 immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 hypotensive; dermatological; immunosuppressive; antineumatic; antihypertensive;  
 antiviral; antibacterial; antifungal; antipneumatic; antihypertensive;  
 antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 cholesterol ester storage; systemic lupus erythematosus; infection;  
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 bone damage; cartilage damage; antinflammatory disease; coagulation;  
 thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0340763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;  
 XX WPI; 2000-602362/57.  
 DR N-PSDB; AAC76820.

PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 3927; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytosolic; hepatotropic; vulnery;  
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressive;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antinflammatory; antibacterial; antiviral; antifungal; antipneumatic;  
 CC antihypertensive; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antinflammatory disease, to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SO Sequence 276 AA;

Query Match 76.5%; Score 26; DB 21; Length 276;  
 Best Local Similarity 75.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVKXVAEF 8  
 II IIII  
 Db 156 EVKDAEF 163

RESULT 14

AAB57116  
 ID AAB57116 standard; Protein; 282 AA.

XX AAB57116;

DT 13-MAR-2001 (first entry)

XX Human prostate cancer antigen protein sequence SEQ ID NO:1694.

Human: prostate cancer; prostate cancer antigen; detection; diagnosis;  
 human; vulnery; cytosolic; hepatotropic; immunomodulatory; muscular;  
 antiproliferative; antiparkinsonian; neurotrophic; antidiabetic;  
 antibacterial; gene therapy; neural; immune; reproductive; renal;  
 gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 wound; infectious disease.

XX Homo sapiens.

XX WO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 : Search time 1.14005 seconds  
(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-81

Perfect score: 34

Sequence: 1 EVMXVAEF 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCtus.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	76.5	559	4	US-09-242-690A-15
2	25	73.5	54	2	US-08-456-647B-18
3	25	73.5	54	2	US-08-237-401A-18
4	25	73.5	405	4	US-09-231-023A-20
5	25	73.5	485	2	US-08-446-803-2
6	25	73.5	485	2	US-08-861-837-2
7	25	73.5	485	3	US-08-600-656-2
8	25	73.5	485	4	US-09-170-670-2
9	25	73.5	485	4	US-09-170-670-8
10	25	73.5	485	4	US-09-193-068-2
11	25	73.5	485	4	US-09-193-068-8
12	25	73.5	485	4	US-09-183-412-2
13	25	73.5	485	4	US-09-183-412-8
14	25	73.5	485	4	US-09-264-097-5
15	25	73.5	485	4	US-09-354-191A-2
16	25	73.5	513	2	US-08-459-346-19
17	25	73.5	513	2	US-07-989-847-8
18	25	73.5	513	3	US-08-889-419-19
19	25	73.5	513	3	US-08-469-411-8
20	25	73.5	513	6	PCT-US93-07189-19
21	25	73.5	513	6	5187076-6
22	25	73.5	556	3	US-08-505-377-1
23	25	73.5	556	3	US-08-798-269-1
24	25	73.5	536	4	US-09-055-210-1
25	25	73.5	816	1	US-07-640-029-1
26	25	73.5	817	1	US-07-640-029-2
27	24	70.6	57	4	US-08-630-915A-209

28	24	70.6	161	4	US-08-858-207A-284	Sequence 284, App
29	24	70.6	329	1	US-08-230-047-7	Sequence 7, Appl
30	24	70.6	341	1	US-08-314-309A-19	Sequence 19, Appl
31	24	70.6	524	3	US-08-557-210A-3	Sequence 3, Appl
32	24	70.6	539	3	US-08-557-210A-5	Sequence 4, Appl
33	24	70.6	539	3	US-08-557-210A-5	Sequence 5, Appl
34	24	70.6	691	5	PCT-US91-08442-2	Sequence 2, Appl
35	24	70.6	758	1	US-07-736-250-16	Sequence 16, Appl
36	24	70.6	972	3	US-08-335-844A-24	Sequence 24, Appl
37	23	67.6	41	1	US-08-096-741-1	Sequence 46, Appl
38	23	67.6	54	2	US-08-456-647B-46	Sequence 47, Appl
39	23	67.6	54	2	US-08-456-647B-47	Sequence 46, Appl
40	23	67.6	54	2	US-08-237-401A-46	Sequence 47, Appl
41	23	67.6	54	2	US-08-237-401A-47	Sequence 46, Appl
42	23	67.6	83	1	US-08-096-741-7	Sequence 47, Appl
43	23	67.6	215	1	US-08-431-080-22	Sequence 22, Appl
44	23	67.6	215	2	US-08-938-534-22	Sequence 22, Appl
45	23	67.6	236	4	US-09-605-858-34	Sequence 34, Appl

## ALIGNMENTS

```
RESULT 1
US-09-242-690A-15
: Sequence 15, Application US/09242690A
: Patent No. 6284534
: GENERAL INFORMATION:
: APPLICANT: MIURA, YUTAKA
: TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
: FILE REFERENCE: 049441/0118
: CURRENT FILING DATE: 1999-02-23
: PRIOR FILING DATE: 1997-08-22
: PRIOR APPLICATION NUMBER: PCT/JP97/02924
: PRIOR FILING DATE: 1996-08-23
: NUMBER OF SEQ ID NOS: 66
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 15
: LENGTH: 559
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic protein
US-09-242-690A-15
Query Match          76.5%; Score 26; DB 4; Length 559;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8
Db 138 EIMPIAOF 145

RESULT 2
US-08-456-647B-18
: Sequence 18, Application US/08456647B
: Patent No. 581516
: GENERAL INFORMATION:
: APPLICANT: Lemke Ph.D. et al., Greg E.
: TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
: NUMBER OF SEQUENCES: 54
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: US
: ZIP: 92037
```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,647B
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/237,401
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/884,486
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07251/007002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-456-647B-18

Query Match          73.5%; Score 25; DB 2; Length 54;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      1 EVMXVAEF 8
      :|:|:|:|
Db      8 DVMKIADE 15

RESULT 3
US-08-237-401A-18
; Sequence 18, Application US/08237401A
; Patent No. 5837448
; GENERAL INFORMATION:
; APPLICANT: Lemke Ph.D. et al., Greg E.
; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,401A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/884,486
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099

```

```

;
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-237-401A-18

Query Match          73.5%; Score 25; DB 2; Length 54;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      1 EVMXVAEF 8
      :|:|:|:|
Db      8 DVMKIADE 15

RESULT 4
US-09-291-023A-20
; Sequence 20, Application US/09291023A
; Patent No. 6309871
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Borchert, Torben
; APPLICANT: Nielsen, Bjarne
; APPLICANT: Nielsen, Videke
; APPLICANT: Hoeck, Lisbeth
; TITLE OF INVENTION: Polypeptides Having Alkaline Alpha-amylase Activity And Nuclei
; FILE REFERENCE: 5821.010-US
; CURRENT APPLICATION NUMBER: US/09/291,023A
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: DK 1999 00438
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 20
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Bacillus
; US-09-291-023A-20

Query Match          73.5%; Score 25; DB 4; Length 405;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      1 EVMXVAEF 8
      :|:|:|:|
Db      180 EFMVAEAF 187

RESULT 5
US-08-446-803-2
; Sequence 2, Application US/08446803
; Patent No. 5824531
; GENERAL INFORMATION:
; APPLICANT: Ottup, Helle
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Rasmussen, Michael Dolberg
; APPLICANT: Van Der Zee, Pia
; TITLE OF INVENTION: Alkaline Bacillus Amylase
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 58245310 No. 5824531disk of No. 5824531th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,803  
FILING DATE: 01-June-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4157.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 867-0123  
TELEFAX: (212) 878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-446-803-2

Query Match 73.5%; Score 25; DB 2; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXKVAEF 8  
I: |||||  
DB 260 EMPVAEAF 267

RESULT 6  
US-08-861-837-2  
Sequence 2, Application US/0861837  
Patent No. 5856164  
GENERAL INFORMATION:  
APPLICANT: Orlup, Helle  
APPLICANT: Bisgard-Frantzen, Henrik  
APPLICANT: Ostergaard, Peter Rahbek  
APPLICANT: Rasmussen, Michael Dolberg  
APPLICANT: Van Der Zee, Pia  
TITLE OF INVENTION: Alkaline Bacillus Amylase  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 58561640 No. 5856164disk of No. 5856164th America  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/861,837  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/446,803  
FILING DATE: 01-June-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4157.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 867-0123  
TELEFAX: (212) 878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-861-837-2

Query Match 73.5%; Score 25; DB 2; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXKVAEF 8  
I: |||||  
DB 260 EMPVAEAF 267

RESULT 7  
US-08-600-656-2  
Sequence 2, Application US/08600656  
Patent No. 6093562  
GENERAL INFORMATION:  
APPLICANT: Bisgard-Frantzen, Henrik  
APPLICANT: Svendsen, Allan  
APPLICANT: Borchert, Torben Vedel  
TITLE OF INVENTION: AMYLASE VARIANTS  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60935620 No. 6093562disk of No. 6093562th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,656  
FILING DATE: 13-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4318.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-600-656-2

Query Match 73.5%; Score 25; DB 3; Length 485;  
Best Local Similarity 62.5%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXKVAEF 8  
I: |||||  
DB 260 EMPVAEAF 267

RESULT 8  
US-09-170-670-2  
Sequence 2, Application US/09170670  
Patent No. 6187576  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Borchert, Torben  
APPLICANT: Bisgard-Frantzen, Henrik

```
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 5276.200-US
; CURRENT APPLICATION NUMBER: US/09/170,670
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 1172/97
; EARLIER FILING DATE: 1997-10-13
; EARLIER APPLICATION NUMBER: 60/063,306
; EARLIER FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-170-670-2

Query Match
Best Local Similarity 73.5%; Score 25; DB 4; Length 485;
Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8
|: ||||
Db 260 EMFAVAEF 267

RESULT 9
US-09-170-670-8
; Sequence 8, Application US/09170670
; Patent No. 6187576
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben
; APPLICANT: Bisgaard-Frantzen, Henrik
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 5276.200-US
; CURRENT APPLICATION NUMBER: US/09/170,670
; CURRENT FILING DATE: 1998-10-13
; EARLIER APPLICATION NUMBER: 1172/97
; EARLIER FILING DATE: 1997-10-13
; EARLIER APPLICATION NUMBER: 60/063,306
; EARLIER FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-170-670-8

Query Match
Best Local Similarity 73.5%; Score 25; DB 4; Length 485;
Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8
|: ||||
Db 260 EMFAVAEF 267

RESULT 10
US-09-193-068-2
; Sequence 2, Application US/09193068
; Patent No. 6197565
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Kjærulff, Søren
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Andersen, Carsten
; TITLE OF INVENTION: -Amylase Variants
; FILE REFERENCE: 5709.000-US
; CURRENT APPLICATION NUMBER: US/09/193,068
; CURRENT FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 2
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-193-068-2

Query Match
Best Local Similarity 73.5%; Score 25; DB 4; Length 485;
Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8
|: ||||
Db 260 EMFAVAEF 267

RESULT 11
US-09-193-068-8
; Sequence 8, Application US/09193068
; Patent No. 6197565
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Kjærulff, Søren
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Andersen, Carsten
; TITLE OF INVENTION: -Amylase Variants
; FILE REFERENCE: 5709.000-US
; CURRENT APPLICATION NUMBER: US/09/193,068
; CURRENT FILING DATE: 1998-11-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-193-068-8

Query Match
Best Local Similarity 73.5%; Score 25; DB 4; Length 485;
Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8
|: ||||
Db 260 EMFAVAEF 267

RESULT 12
US-09-183-412-2
; Sequence 2, Application US/09183412
; Patent No. 6204232
; GENERAL INFORMATION:
; APPLICANT: Borchert, Torben V.
; APPLICANT: Svendsen, Allan
; APPLICANT: Andersen, Carsten
; APPLICANT: Nielsen, Bjarne
; APPLICANT: Nissen, Torben L.
; APPLICANT: Kjærulff, Søren
; TITLE OF INVENTION: Alpha-Amylase Mutants
; FILE REFERENCE: 5368.200-US
; CURRENT APPLICATION NUMBER: US/09/183,412
; CURRENT FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: 60/064,662
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 60/093,234
; EARLIER FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 1240/97
; EARLIER FILING DATE: 1997-10-30
; EARLIER APPLICATION NUMBER: PA 1998 00936
; EARLIER FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 485
; TYPE: PRT
```

```

; ORGANISM: Bacillus sp.
US-09-183-412-2
Query Match          73.5%: Score 25; DB 4; Length 485;
Best Local Similarity 62.5%: Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXVAEF 8
   1: ||||
Db 260 EMFAVAEF 267

RESULT 13
US-09-183-412-8
; Sequence 8, Application US/09183412
; Patent No. 6204232
; GENERAL INFORMATION:
; APPLICANT: Borchert, Torben V.
; APPLICANT: Svendsen, Allan
; APPLICANT: Andersen, Carsten
; APPLICANT: Nielsen, Bjarne
; APPLICANT: Nielsen, Torben L.
; APPLICANT: Kjaerulf, Soren
; TITLE OF INVENTION: Alpha-Amulase Mutants
; FILE REFERENCE: 5368 200-US
; CURRENT APPLICATION NUMBER: US/09/183,412
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: 60/064,662
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 60/093,234
; EARLIER FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 1240/97
; EARLIER FILING DATE: 1997-10-30
; EARLIER APPLICATION NUMBER: PA 1998 00936
; EARLIER FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-183-412-8

Query Match          73.5%: Score 25; DB 4; Length 485;
Best Local Similarity 62.5%: Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXVAEF 8
   1: ||||
Db 260 EMFAVAEF 267

RESULT 14
US-09-264-097-5
; Sequence 5, Application US/09264097
; Patent No. 6287826
; GENERAL INFORMATION:
; APPLICANT: No. 6287826man; Barrie Edmund
; APPLICANT: Hendriksen, Hanne Vang
; TITLE OF INVENTION: Enzymatic Preparation of Glucose Syrup
; FILE REFERENCE: 5278. 200-US
; CURRENT APPLICATION NUMBER: US/09/264,097
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: PA 0321/98
; EARLIER FILING DATE: 1998-03-09
; EARLIER APPLICATION NUMBER: 60/079,209
; EARLIER FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 485
; TYPE: PRT
```

```

; ORGANISM: Bacillus
US-09-264-097-5
Query Match          73.5%: Score 25; DB 4; Length 485;
Best Local Similarity 62.5%: Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXVAEF 8
   1: ||||
Db 260 EMFAVAEF 267

RESULT 15
US-09-354-191A-2
; Sequence 2, Application US/09354191A
; Patent No. 6297038
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; APPLICANT: Borchert, Torben Vedel
; TITLE OF INVENTION: AMYLASE VARIANTS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 62970380 No. 6297038disk of No. 6297038th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/354,191A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/600,656
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4318.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-354-191A-2

Query Match          73.5%: Score 25; DB 4; Length 485;
Best Local Similarity 62.5%: Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVXVAEF 8
   1: ||||
Db 260 EMFAVAEF 267

Search completed: October 30, 2002, 12:32:34
Job time : 2.14005 secs
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002, Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 1.5317 Seconds  
(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-81

Sequence: 1 EWMXVAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	85.3	195	2	C97255
2	29	85.3	236	2	B70728
3	28	82.4	594	2	A10673
4	28	82.4	1256	2	AB2042
5	28	82.4	2638	1	A42545
6	27	79.4	181	2	T11902
7	27	79.4	184	2	C69133
8	27	79.4	201	2	F69988
9	27	79.4	598	2	B71095
10	27	79.4	601	2	D96001
11	27	79.4	802	2	T05596
12	27	79.4	1560	2	T30282
13	26	76.5	46	2	B55209
14	26	76.5	150	2	B55209
15	26	76.5	223	2	B64205
16	26	76.5	224	2	S02216
17	26	76.5	252	2	E72060
18	26	76.5	252	2	B86564
19	26	76.5	269	2	A56487
20	26	76.5	278	2	C66421
21	26	76.5	294	2	T15662
22	26	76.5	324	2	B84452
23	26	76.5	490	2	E96010
24	26	76.5	498	2	H82494
25	26	76.5	533	2	A61616
26	26	76.5	533	2	A61253
27	26	76.5	543	2	G83825
28	26	76.5	558	2	J05135
29	26	76.5	561	2	S73087

30	26	76.5	562	2	G75044	acetoacetate synth
31	26	76.5	587	2	D81881	probable succinate
32	26	76.5	587	2	F81138	succinate dehydrog
33	26	76.5	603	2	D70445	aspartate--tRNA li
34	26	76.5	623	2	B82536	ABC transporter At
35	26	76.5	1014	2	C83990	beta-galactosidase
36	26	76.5	1085	2	G89056	protein K09H11.3 l
37	25	73.5	136	2	I70179	dyad protein - Esc
38	25	73.5	152	2	S14388	hypothetical prote
39	25	73.5	152	2	E90639	conserved hypotnet
40	25	73.5	152	2	E85490	conserved hypotnet
41	25	73.5	152	2	AC0067	acetyltransferase [i
42	25	73.5	152	2	AD0517	dynein-like protei
43	25	73.5	187	2	A86847	
44	25	73.5	188	2	I70180	
45	25	73.5	200	2	I70169	

#### ALIGNMENTS

RESULT 1  
C97255  
thymidine kinase (EC 2.7.1.21) [similarity] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 31-Dec-2001  
C:Accession: C97255  
R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A86900; MUID:21359325; PMID:21359325  
A:Accession: C97255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AKR80830.1; PID:G15025935; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2887  
C:Superfamily: thymidine kinase  
C:Keywords: phosphotransferase

Query Match 85.3% Score 29; DB 2; Length 195;  
Best Local Similarity 62.5%; Pred. No. 9;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EWMXVAEF 8  
DB 130 ELMXVAEF 137

RESULT 2  
B70728  
hypothetical protein RV2558 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1998  
C:Accession: B70728  
R:Conor, R.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Rajadream, M.A.; Rogers, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:9825987  
A:Accession: B70728  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-236 <COL>  
A:Cross-references: GB:277250; GB:AL12456; NID:93261617; PIDN:CA001046.1; PID:e25533  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV2558

Query Match 85.3%; Score 29; DB 2; Length 236;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVXVAF 8  
 ||| |||  
 DB 216 EVLDAF 223

## RESULT 3

A10673  
 probable hydrolase STY1503 [imported] - Salmonella enterica subsp. enterica serovar Typh  
 C:Species: Salmonella enterica subsp. enterica serovar Typh  
 A:Note: this species has also been called Salmonella typh  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: A10673  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typh  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: A10673  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-594 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD01762.1; PID:916502610; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY1503  
 C:Superfamily: trehalose trehalohydrolase

Query Match 82.4%; Score 28; DB 2; Length 594;  
 Best Local Similarity 75.0%; Pred. No. 52;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVXVAF 8  
 ||| |||  
 DB 148 EVMPVAF 155

## RESULT 4

AB2042  
 hypothetical protein all1888 [imported] - Anabaena sp. (strain PCC 7120)  
 C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: AB2042  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; M01D:21595285; PMID:1175840  
 A:Accession: AB2042  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1256 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA073587.1; PID:g17130978; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all1888

Query Match 82.4%; Score 28; DB 2; Length 1256;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVXVAF 8  
 ||| |||  
 DB 74 EVMAF 81

## RESULT 5

A42545  
 genome polypeptide - Langat virus (strain TP21) (fragment)  
 N:Contains: nonstructural protein NS1; nonstructural protein NS2a; nonstructural protein NS5

C:Species: Langat virus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Jan-2001  
 C:Accession: A42545; A61409; C61409  
 R:Jacono-Connors, L.C.; Schmaljohn, C.S.  
 Virology 188, 875-880, 1992  
 A:Title: Cloning and sequence analysis of the genes encoding the nonstructural proteins NS1 and NS2a of the Langat virus  
 A:Reference number: A42545; M01D:92263794  
 A:Accession: A42545

A:Molecule type: genomic RNA  
 A:Residues: 1-2638 <IAC>  
 A:Cross-references: GB:S53565; NID:g249315; PIDN:AA022165.1; PID:g249316  
 R:Gulikho, F.; Heinz, F.X.; Mandl, C.W.; Holzmann, H.; Kunz, C.; Gresslko, M.  
 J. Gen. Virol. 72, 333-338, 1991  
 A:Title: The relationship between the flaviviruses Skalka and Langat as revealed by  
 A:Reference number: A61409; M01D:91132129  
 A:Accession: A61409  
 A:Status: not compared with conceptual translation  
 A:Molecule type: genomic RNA  
 A:Residues: 319-337 <GUT>  
 A:Accession: C61409  
 A:Status: not compared with conceptual translation  
 A:Molecule type: genomic RNA  
 A:Residues: 877-994 <GUT>  
 C:Superfamily: yellow fever virus genome polypeptide  
 C:Keywords: glycoprotein; nonstructural protein; nucleotide binding; P-loop; polypeptide  
 F:1-352/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1-352/Product: nonstructural protein NS2a #status predicted <NS2>  
 F:353-382/Product: nonstructural protein NS2b #status predicted <NS2>  
 F:583-713/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:714-1334/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:912-919/Region: nucleotide-binding motif A (P-loop)  
 F:1335-1483/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1484-1735/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:1736-2638/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:85,207,223,873,1212,1671,1950/Binding site: carbohydrate (Asn) #status p

Query Match 82.4%; Score 28; DB 1; Length 2638;  
 Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 VMXVAF 8  
 ||| |||  
 DB 153 VMTVAF 159

## RESULT 6

T11902  
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain NQ02 - Thermus aquaticus thermophilus  
 C:Species: Thermus aquaticus thermophilus  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 C:Accession: T11902  
 R:Yano, T.; Chu, S.S.; Sled, V.S.; Ohnishi, T.; Yagi, T.  
 submitted to: The EMBL Data Library, March 1996  
 A:Description: Cloning, sequencing, and expression studies of the proton-translocating  
 A:Reference number: Z17372  
 A:Accession: T11902  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-181 <YAN>  
 A:Cross-references: EMBL:U52917; NID:g1279860; PID:g1279865; PIDN:AAA97942.1  
 C:Genetics:  
 A:Gene: NQ02  
 C:Keywords: electron transfer; membrane-associated complex; NAD; oxidoreductase

Query Match 79.4%; Score 27; DB 2; Length 181;  
 Best Local Similarity 75.0%; Pred. No. 27;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVXVAF 8  
 ||| |||

Db 59 EVWGVASF 66

## RESULT 7

C69133

DNA-dependent RNA polymerase, subunit E' - Methanobacterium thermoautotrophicum (strain C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: C69133

R:Smith, D.R.; Doucette-Stamm, L.A.; Delouche, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Olin, D.; Spadafora, R.; Vicalini, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.;  
K., S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
A:Reference number: A69000; MUID:98037514  
A:Accession: C69133

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-184 <MTH>

A:Cross-references: GB:AE000812; GB:AE000666; NID:g2621298; PIDN:AAB84770.1; PID:g262131  
A:Experimental source: strain Delta H  
C:Genetics:

A:Gene: MTH264  
A:Start codon: TTG  
C:Superfamily: DNA-directed RNA polymerase subunit E

Query Match Best Local Similarity 79.4%; Score 27; DB 2; Length 184;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVWVAAF 8  
11: |||  
Db 88 EVIRIAEF 95

## RESULT 8

F69988

hypothetical protein ytbQ - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: F69988

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Berte  
C.: Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Ch  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall  
lech, J.; Harwood, C.R.; Hennut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koestler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
Y, M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scallion,  
A:Authors: Schleich, S.; Schoefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
A:Authors: Schleich, S.; Schoefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
T.; Waters, P.; Wipat, A.; Yamamoto, H.; Yamane, E.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Dancin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033  
A:Accession: F69988

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-201 <KUN>  
A:Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14996.1; PID:el185891;  
A:Experimental source: strain 168  
C:Genetics:

A:Gene: ytbQ

Query Match Best Local Similarity 79.4%; Score 27; DB 2; Length 201;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVWVAAF 8  
11: |||  
Db 22 DIMDIAEF 29

## RESULT 9

B71095

hypothetical protein PH1023 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C:Accession: B71095

R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Se  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Og  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil  
A:Reference number: A71000; MUID:98344137  
A:Accession: B71095

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-598 <KAW>  
A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30120.1; PID:g3257437  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by Genba  
C:Genetics:

A:Gene: PH1023

Query Match Best Local Similarity 79.4%; Score 27; DB 2; Length 598;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVWVAAF 8  
11: |||  
Db 280 EIVNVAEF 287

## RESULT 10

D96001

Probable 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) [imported] - Sinorhizobium m  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001  
C:Accession: D96001

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing e  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: D96001  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-601 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49676.1; PID:g15141163; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymb  
R:Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
hebut, P.; Vandemol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:

A:Gene: glgB2; SMB21447  
A:Gene: plasmid  
C:Superfamily: trehalose trehalohydrolase  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match Best Local Similarity 79.4%; Score 27; DB 2; Length 601;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVWVAAF 8  
11: |||  
Db 151 EIMVPAQF 158

## RESULT 11

T05596  
 Probable potassium transport protein F9D16.110 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 18-Aug-2000  
 C:Accession: T05596  
 R:Bayan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohenisel, J.; Mewes, H.W.; Meyer, K.F.  
 Submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15419  
 A:Accession: T05596  
 A:Molecule type: DNA  
 A:Residues: 1-802 <BEV>  
 A:Cross-references: EMBL:AL035394  
 A:Experimental source: cultivar Columbia; BAC clone F9D16  
 A:Genetics:  
 C:Map position: 4  
 A:Introns: 1/3; 16/3; 102/1; 185/1; 209/1; 296/1; 313/3; 352/1; 437/1  
 A:Note: F9D16.110  
 C:Superfamily: barley probable potassium transport protein HAK1  
 C:Keywords: ion transport

Query Match 79.4%; Score 27; DB 2; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VMXVAEF 8  
 ||: |||  
 Db 629 VMSIAEF 635

RESULT 12  
 T30282  
 calcium-binding protein - sea urchin (Strongylocentrotus purpuratus)  
 C:Species: Strongylocentrotus purpuratus (purple urchin)  
 C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T30282  
 R:Solystik-Espanola, M.; Kilmzing, D.C.; Pfarr, K.; Burke, R.D.; Ernst, S.G.  
 Dev. Biol. 165, 73-85, 1994  
 A:Title: Endo16, a large multidomain protein found on the surface and ECM of endodermal  
 A:Reference number: Z20805; MUID:94374583  
 A:Accession: T30282  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 11560 <SC0>  
 A:Cross-references: EMBL:L34680; NID:g511893; PID:g511894; PIDN:AAA30047.1  
 C:Genetics:  
 A:Note: endo16

Query Match 79.4%; Score 27; DB 2; Length 1560;  
 Best Local Similarity 62.5%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8  
 ||: |||  
 Db 705 ELAAVAEF 712

RESULT 13  
 F95318  
 conserved hypothetical protein Sma0833 [imported] - Sinorhizobium meliloti (strain 1021)  
 C:Species: Sinorhizobium meliloti  
 C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: F95318  
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
 ; Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: F95318  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-46 <KUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK5112.1; PID:g14523550; GSPDB:GN00165  
 A:Experimental source: strain 1021, megaplasmid pSyma

R:Gallibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
 ela, D.; Chah, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeJau  
 hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Sma0833  
 A:Genome: plasmid

Query Match 76.5%; Score 26; DB 2; Length 46;  
 Best Local Similarity 71.4%; Pred. No. 11;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAE 7  
 ||: |||  
 Db 17 EIMAVAE 23

RESULT 14  
 B55209  
 H transfer determinant A - plasmid R27  
 C:Species: plasmid R27  
 C>Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 05-Nov-1999  
 C:Accession: B55209  
 R:Whelan, K.F.; Maher, D.; Colleran, E.; Taylor, D.E.  
 J. Bacteriol. 176, 2242-2251, 1994  
 A:Title: Genetic and nucleotide sequence analysis of the gene htdA, which regulates c  
 A:Reference number: A55209; MUID:94209223  
 A:Accession: B55209  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-150 <WHE>  
 A:Cross-references: GB:L20342; NID:g410304; PIDN:AAB00506.1; PID:g1326032  
 C:Genetics:  
 A:Gene: htdA  
 A:Genome: plasmid

Query Match 76.5%; Score 26; DB 2; Length 150;  
 Best Local Similarity 50.0%; Pred. No. 39;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8  
 ||: |||  
 Db 18 EVLSISEF 25

RESULT 15  
 E64205  
 deoxyribose-phosphate aldolase (EC 4.1.2.4) - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C>Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 21-Jul-2000  
 C:Accession: E64205  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.  
 M.; Fuhrmann, J.; Nguyen, D.; Uitterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,  
 C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346  
 A:Accession: E64205  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-223 <TIG>  
 A:Cross-references: GB:U03684; GB:LA3967; NID:g38444650; PIDN:AACT1266.1; PID:g1045723  
 A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: GCG3  
 C:Superfamily: deoxyribose-phosphate aldolase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 76.58; Score 26; DB 2; Length 223;  
Best Local Similarity 71.48; Pred. No. 59;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 VMAXAEF 8  
11:111  
DB 93 VMAXAEF 99

Search completed: October 30, 2002, 12:31:34  
Job time : 4:53:17 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 30, 2002, 12:21:10 ; Search time 0.687961 Seconds  
(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-81

Perfect score: 34

Sequence: 1 EYMXVAEP 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	29	85.3	236	1 YP58_MYCTU	Q50740 mycobacteri
2	28	82.4	3414	1 POLG_LANTY	P28837 l genome po
3	27	79.4	181	1 N002_THETR	O56221 thermus aqu
4	27	79.4	253	1 YRBO_BACSU	P33560 bacillus su
5	26	76.5	223	1 DEOC_MYCE	P47296 mycoplasma
6	26	76.5	224	1 DEOC_MYCPN	P09924 mycoplasma
7	26	76.5	269	1 SRPB_MOUSE	P47558 mus musculu
8	26	76.5	271	1 SRPB_HUMAN	O95588 homo sapien
9	26	76.5	457	1 ARLY_PASMC	P57909 pasteurilla
10	26	76.5	603	1 SYD_AOUAE	O67589 aquilex aeo
11	25	73.5	152	1 YABB_ECOLI	P22186 escherichia
12	25	73.5	413	1 ZABB_RABIT	O00006 o serine/th
13	25	73.5	426	1 ZABA_PIG	O29090 s serine/th
14	25	73.5	443	1 ZABB_HUMAN	O00005 homo sapien
15	25	73.5	443	1 ZABB_PIG	P54614 sus scrofa
16	25	73.5	447	1 ZABB_RAT	P36877 r serine/th
17	25	73.5	447	1 ZABA_HUMAN	Q00007 h serine/th
18	25	73.5	447	1 ZABA_RAT	P36876 r serine/th
19	25	73.5	453	1 ZABD_RAT	P56932 r serine/th
20	25	73.5	499	1 ZABA_DROME	P36872 drosophila
21	25	73.5	513	1 BWP6_HUMAN	P32004 homo sapien
22	25	73.5	632	1 EFPD_SCHPO	P87111 s probable
23	25	73.5	808	1 EGR4_MOUSE	O03142 mus musculu
24	25	73.5	4466	1 DTHC_ANTCR	P39057 antiochidari
25	25	73.5	4466	1 DTHC_TRIGGER	P33098 tripeustes
26	25	73.5	4486	1 DVHO_HUMAN	O9569 homo sapien
27	24	70.6	91	1 YAHQ_ECOLI	P75694 escherichia
28	24	70.6	145	1 RL13_HALMA	P29189 halorcula
29	24	70.6	148	1 CYC6_CHIRE	P68189 chlamydomon
30	24	70.6	157	1 RAP_TAROF	O49065 taraxacum o
31	24	70.6	186	1 RST_METJA	O27130 methanobact
32	24	70.6	187	1 REP1_METJA	O57840 methanococc
33	24	70.6	263	1 NIH2_METIV	P08624 methanobact

34	24	70.6	291	1	AMPN_ARCFU	O28438 archaeoglob
35	24	70.6	295	1	STOE_BOVIN	P19217 bos taurus
36	24	70.6	298	1	YSMK_CAEL	O19408 caenorhabdi
37	24	70.6	341	1	KHYB_ECOLI	P00557 escherichia
38	24	70.6	377	1	NCK1_HUMAN	P16333 homo sapien
39	24	70.6	402	1	OPS4_CANAL	P46596 candida alb
40	24	70.6	431	1	PURA_EDWIC	O31047 edwardsiell
41	24	70.6	474	1	YPC2_CAEL	O11179 caenorhabdi
42	24	70.6	549	1	TCPA_CAEL	P41988 caenorhabdi
43	24	70.6	565	1	TP6B_AERPE	O9564 aeropyrum p
44	24	70.6	758	1	VKGC_BOVIN	O07175 bos taurus
45	24	70.6	758	1	VKGC_HUMAN	P38435 homo sapien

## ALIGNMENTS

RESULT 1  
ID YP58\_MYCTU STANDARD: PRT: 236 AA.  
AC Q50740:

DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 25.7 kDa protein RV2558.  
GN RV2558 OR MT2635 OR MTCY9C4.10C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
NCBI\_TaxID=1773;  
[1]  
SEQUENCE FROM N.A.

RA MEDLINE=96295987; PubMed=6634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Stulson J.E., Taylor K., Whitehead S., Barrrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
[2]

RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
-i- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2557.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----

DR EMBL: Z77250; CAB01046.1; -  
DR EMBL: AE007098; AAK46947.1; -  
DR TIGR: MT2635; -  
DR Tuberculist: RV2558; -  
KW Hypothetical protein: Complete proteome.  
SQ SEQUENCE 236 AA; 25718 MW; 13E3B049DBF79C6B CRC64;





```

GN  N002.
OS  Thermus aquaticus (subsp. thermophilus).
CC  Bacteria; Thermus/Denitococcus group; Thermus group; Thermus.
OX  NCBI_TaxID=274;
RN  [1]
RP  SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC  STRAIN=HB8 / ATCC 27634;
RX  MEDLINE=97172490; PubMed=9020134;
RA  Yano T., Chu S.S., Sled V.D., Ohnishi T., Yagi T.;
RT  "The proton-translocating NADH-quinone oxidoreductase (NDH-1) of
RT  thermophilic bacterium Thermus thermophilus HB-8. Complete DNA
RT  sequence of the gene cluster and thermostable properties of the
RT  expressed N002 subunit."
RL  J. Biol. Chem. 272:4201-4211(1997).
CC  -1- CATALYTIC ACTIVITY: NADH + ubiquinol = NAD(+) + ubiquinol.
CC  -1- COFACTOR: BINDS A 2FE-2S CLUSTER.
CC  -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.
CC  -1- SIMILARITY: BELONGS TO THE COMPLEX I 24 KDA SUBUNIT FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC  or send an email to license@sdb.ch.)
CC  -----
DR  EMBL: U52917; AAA97942.1; -
DR  InterPro: IPR002023; Complex1_24KD.
DR  Pfam: PF01257; complex1_24KD; 1.
DR  ProDom: PD003859; Complex1_24KD; 1.
DR  ProSITE: PS01099; Complex1_24K; 1.
KM  Oxidoreductase; NAD; ubiquinol; Iron-sulfur.
FT  METAL 83 83 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 88 88 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 124 124 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 128 128 IRON-SULFUR (2FE-2S) (POTENTIAL).
SQ  SEQUENCE 181 AA; 20286 MW; 484FED9245C613EE CRC64;

Query Match
Best Local Similarity 79.4%; Score 27; DB 1; Length 181;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY  1 EVXVXAEF 8
    |||||
DB  59 EVMGVASF 66

RESULT 4
YTBO_BACSU STANDARD: PRT: 253 AA.
AC P53560:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ytbQ.
GN YTBO.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96312354; PubMed=8763340;
RA Bower S., Perkins J.B., Vocum R.R., Howitt C.L., Rahaim P.,
RA Pero J.;
RT "Cloning, sequencing, and characterization of the Bacillus subtilis
RT biotin biosynthetic operon."
RL J. Bacteriol. 178:4122-4130(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch.)
CC -----
DR EMBL: U51868; AAB17463.1; -
DR Subtilist; BG11787; ytbQ.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 28233 MW; 572CC4681F04C650 CRC64;

Query Match
Best Local Similarity 79.4%; Score 27; DB 1; Length 253;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY  1 EVXVXAEF 8
    |||||
DB  74 DIMDIAEF 81

RESULT 5
DEOC_MYCGE STANDARD: PRT: 223 AA.
AC P47296;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
DE (Deoxyriboaldolase).
GN DEOC OR MG050.
OS Mycoplasma genitalium.
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
CC Mycoplasmales; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Nguyen D.T., Ullrich T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Luetter T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC glyceraldehyde 3-phosphate + acetaldehyde.
CC -1- PATHWAY: NUCLEOTIDE AND DEOXYRIBONUCLEOTIDE CATABOLISM.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.
CC DEOC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch.)
CC -----
DR EMBL: U39684; AAC71266.1; -
DR TIGR: MG050; -
DR InterPro: IPR002915; Deoc.
DR InterPro: IPR003009; FMN_enzyme.
DR Pfam: PF01791; Deoc; 1.
KM Lyase; Schiff base; Complete proteome.
FT BINDING 152 152 SCHIFF-BASE (BY SIMILARITY).
SQ SEQUENCE 223 AA; 24675 MW; 33243023ICE99DB0 CRC64;

Query Match
Best Local Similarity 76.5%; Score 26; DB 1; Length 223;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

OY 2 VMXVAEF 8  
11 :111  
DB 93 VMNTAEF 99

RESULT 6  
DEOC\_MYCPN STANDARD; PRT; 224 AA.

AC P09924;  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)  
DE (Deoxyriboaldolase).  
GN DEOC OR MPN053 OR MP091.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;

SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=89128453; PubMed=2492658;  
RA Loebel S., Inamine J.M., Hu P.-C.;  
RT "Nucleotide sequence of the deoc gene of Mycoplasma pneumoniae.";  
RL Nucleic Acids Res. 17:801-801(1989).  
[2]

SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelfeich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Hermann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).

CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-  
glyceraldehyde 3-phosphate + acetaldehyde.  
CC -1- PATHWAY: NUCLEOTIDE AND DEOXYRIBONUCLEOTIDE CATABOLISM.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.  
CC DEOC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X13544; CA31897.1; -;  
DR EMBL: AE000011; AAB95739.1; -;  
DR PIR: S02216; S02216.  
DR InterPro: IPR002915; Deoc.  
DR Pfam: PF01791; Deoc. 1.  
KM Lyase; Schiff base; Complete proteome.  
FT BINDING 152 152 SCHIFF-BASE (BY SIMILARITY).  
FT SEQUENCE 224 AA; 24878 MW; 73C3E4932E7881F7 CRC64;

Query Match 76.5%; Score 26; DB 1; Length 224;  
Best Local Similarity 71.4%; Pred. No. 24;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VMXVAEF 8  
11 :111  
DB 93 VMNTAEF 99

RESULT 7  
SRPB\_MOUSE STANDARD; PRT; 269 AA.  
ID SRPB\_MOUSE  
AC P47758; Q9D872;  
DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Signal recognition particle receptor beta subunit (SR-beta).  
GN SRPB.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

SEQUENCE FROM N.A.  
RX MEDLINE=95146535; PubMed=7844142;  
RA Miller J.D., Tajima S., Lauffer L., Walter P.;  
RT "The beta subunit of the signal recognition particle receptor is a  
RT transmembrane GTPase that anchors the alpha subunit, a peripheral  
RT membrane GTPase, to the endoplasmic reticulum membrane.";  
RL J. Cell Biol. 128:273-282(1995).  
[2]

SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Small intestine;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,  
RA Blake J., Boffelli D., Bojunga N., Carinici P., de Bonaldo M.F.,  
RA Brownstein M.J., Bolt C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli R., Momberts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,  
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: HAS GTPASE ACTIVITY. MAY MEDIANE THE MEMBRANE  
CC ASSOCIATION OF SR ALPHA.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.  
CC -1- SIMILARITY: TO OTHER SRP BETA SUBUNITS; DISTANTLY RELATED TO RAS  
CC SUPERFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U17343; AAB69976.1; -;  
DR EMBL: AK008383; AAB2638.1; -;  
DR MGD: MGI:102964; Strpb.  
DR InterPro: IPR003575; Small\_GTPase.  
DR SMART: SM00010; small\_GTPase; 1.  
KM Signal recognition particle; Transmembrane; Receptor;  
KW Endoplasmic reticulum; GTP-binding.  
FT TRANSMEM 35 55 POTENTIAL.  
FT NP\_BIND 69 76 GTP (POTENTIAL).  
FT NP\_BIND 115 119 GTP (POTENTIAL).  
FT NP\_BIND 178 181 GTP (POTENTIAL).  
FT CONFLICT 16 16 A -> P (IN REF. 2).  
FT CONFLICT 173 173 L -> P (IN REF. 1; AA SEQUENCE).  
FT SEQUENCE 269 AA; 29579 MW; 041175FA6891DA37 CRC64;

Query Match 76.5%; Score 26; DB 1; Length 269;  
Best Local Similarity 75.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



```

RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +
CC diphosphate + L-aspareryl-tRNA(Asp).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000750; AAC07548.1; -.
DR HSSP: P36419; 1EFW.
DR InterPro: IPR002106; AA_tRNA_ligase_II.
DR InterPro: IPR004115; GAD.
DR InterPro: IPR002309; tRNA-synt_2.
DR InterPro: IPR002312; tRNA-synt_asep.
DR Pfam: PF02938; GAD; 1.
DR Pfam: PF00152; tRNA-synt_2; 2.
DR Pfam: PF01336; tRNA-anti; 1.
DR PRINTS: PR01042; TRNASYNTHASP.
DR PROSITE: PS00179; AA_tRNA_LIGASE_II_1; 1.
DR PROSITE: PS00339; AA_tRNA_LIGASE_II_2; 1.
DR AMINOACYL-tRNA synthetase; protein biosynthesis; ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 603 AA; 69729 MW; 9DFFBD840C8DCC1C CRC64;

Query Match 76.5%; Score 26; DB 1; Length 603;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYMYVAE 7
Db 254 EYMDVAE 260

RESULT 11
YABB.ECOLI STANDARD; PRT; 152 AA.
ID YABB.ECOLI
AC P22186;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yabb.
GN YABB OR B0081.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90251464; PubMed=2187182;
RA Gomez M.-J., Flooret B., van Heijenoort J., Ayala J.A.;
RT "Nucleotide sequence of the regulatory region of the gene pbbp of
RT Escherichia coli."
RL Nucleic Acids Res. 18:2813-2813(1990).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: BELONGS TO THE UPF0040 FAMILY. STRONG, TO H.INFLUENZAE

```

```

CC H1129.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52063; CA36284.1; -.
DR EMBL: X55034; CA38858.1; -.
DR EMBL: AE000118; AAC73192.1; -.
DR PIR: S14388; S14388.
DR EcoGene: EG11084; yabb.
DR InterPro: IPR003444; UPF0040.
DR Pfam: PF02381; UPF0040; 2.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 152 AA; 17386 MW; 3EE1A6FA9D2B1C01 CRC64;

Query Match 73.5%; Score 25; DB 1; Length 152;
Best Local Similarity 62.5%; Pred. No. 30;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EYMYVAE 8
Db 107 EYMLVGF 114

RESULT 12
YABB.RABIT STANDARD; PRT; 413 AA.
ID YABB.RABIT
AC Q00006;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,
DE beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B,
DE R2-beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,
DE PPR2RB.
GN Oryctolagus cuniculus (Rabbit).
OS Oryctolagus cuniculus; Chordata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=91198016; PubMed=1849734;
RA Mayer R.E., Hendrix P., Cron P., Mathies R., Stone S.R.,
RA Goris J., Merlevede P., Horstenge J., Hemmings B.A.;
RT "Structure of the 55-kDa regulatory subunit of protein phosphatase
RT 2A: evidence for a neuronal-specific isoform."
RL Biochemistry 30:3589-3597(1991).
CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE
CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE
CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR
CC COMPARTMENT.
CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,
CC COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa
CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATE
CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE
CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNTS
CC B (THE R2/B/PR55/B55, R3/B-/PR72/PR130/PR59 AND R5/B-/B56
CC FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,
CC AND CELL SIGNALING MOLECULES.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M64931; AAA31458.1; -  
 DR InterPro: IPR000009; PP2A\_PR55.  
 DR Pfam: PF00400; WD40; 5.  
 DR SMART: SM00320; WD40; 2.  
 DR PROSITE: PS01024; PR55\_1; 1.  
 DR PROSITE: PS01025; PR55\_2; 1.  
 DR Multigene family.  
 DR NON\_TER 1.  
 DR SEQUENCE 413 AA; 48243 MW; 4923787817EB8FE2 CRC64;

Query Match 73.5%; Score 25; DB 1; Length 413;  
 Best Local Similarity 62.5%; Pred. No. 80;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVKXVAEF 8  
 ||: |||  
 Db 193 EVITAAEF 200

RESULT 13  
 2ABA\_PIG STANDARD; PRT; 426 AA.  
 AC Q29050;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,  
 DE alpha isoform (PP2A, subunit B, B-alpha isoform) (PP2A, subunit B,  
 DE B55-alpha isoform) (PP2A, subunit B, PR55-alpha isoform) (PP2A,  
 DE subunit B, R2-alpha isoform) (Fragment).  
 GN PP2A2A.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 NC NCB1\_TaxID=9823;  
 RN [1]  
 RA SEQUENCE FROM R.N.A.  
 RA Mayer-Jaekel R.E.;  
 RL Theiss (1992), Friedrich Miescher Institut / Basel, Switzerland.  
 CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE  
 CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE  
 CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR  
 CC COMPARTMENT.  
 CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,  
 CC COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa  
 CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES  
 CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE  
 CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS  
 CC B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56  
 CC FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,  
 CC AND CELL SIGNALING MOLECULES.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B  
 CC FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z34932; GA84404.1; -  
 DR InterPro: IPR000009; PP2A\_PR55.  
 DR Pfam: PF00400; WD40; 5.  
 DR PRINTS: PR00600; PP2APR55.  
 DR SMART: SM00320; WD40; 3.

DR PROSITE: PS01024; PR55\_1; 1.  
 DR PROSITE: PS01025; PR55\_2; 1.  
 DR Multigene family.  
 DR NON\_TER 1.  
 DR SEQUENCE 426 AA; 49613 MW; 3AAND7EB38B03534 CRC64;

Query Match 73.5%; Score 25; DB 1; Length 426;  
 Best Local Similarity 62.5%; Pred. No. 82;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVKXVAEF 8  
 ||: |||  
 Db 206 EVITAAEF 213

RESULT 14  
 2ABE\_HUMAN STANDARD; PRT; 443 AA.  
 AC Q00005;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,  
 DE beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B,  
 DE beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,  
 DE R2-beta isoform).  
 GN PP2AR2B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA TISSUE=Retal brain;  
 RX MEDLINE=91196016; PubMed=1849734;  
 RA Mayer R.E., Hendrix P., Cron P., Matthies R., Stone S.R.,  
 RA Goris J., Merlevede W., Hofsteenge J., Hemmings B.A.;  
 RT 2A: evidence for a neuronal-specific isoform.;  
 RL Biochemistry 30:3589-3597(1991).  
 CC -1- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE  
 CC SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE  
 CC LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR  
 CC COMPARTMENT.  
 CC -1- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,  
 CC COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa  
 CC CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES  
 CC WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE  
 CC WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS  
 CC B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56  
 CC FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,  
 CC AND CELL SIGNALING MOLECULES.  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B  
 CC FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M64930; AAA36493.1; -  
 DR PIR: B38351; B38351.  
 DR MIM: 604325; -  
 DR InterPro: IPR000009; PP2A\_PR55.  
 DR Pfam: PF00400; WD40; 6.  
 DR PRINTS: PR00600; PP2APR55.  
 DR SMART: SM00320; WD40; 3.  
 DR PROSITE: PS01024; PR55\_1; 1.  
 DR PROSITE: PS01025; PR55\_2; 1.

KW Multigene family.  
SQ SEQUENCE 443 AA; 51710 MW; C383C834B2852B8F CRC64;

Query Match 73.5%; Score 25; DB 1; Length 443;  
Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAF 8  
||: |||  
DB 223 EYTAAEF 230

## RESULT 15

ID 2ABB\_PIG STANDARD; PRT; 443 AA.  
AC P54614;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Serine/threonine protein phosphatase 2A, 55 kDa regulatory subunit B,  
beta isoform (PP2A, subunit B, B-beta isoform) (PP2A, subunit B,  
beta isoform) (PP2A, subunit B, PR55-beta isoform) (PP2A, subunit B,  
beta isoform).  
DE R2-beta isoform).  
GN PP2R2B.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mayer-Jaekel R.E.;  
RL Thesis (1992); Friedrich Miescher Institut / Basel, Switzerland.  
CC -I- FUNCTION: THE B REGULATORY SUBUNIT MIGHT MODULATE SUBSTRATE  
SELECTIVITY AND CATALYTIC ACTIVITY, AND ALSO MIGHT DIRECT THE  
LOCALIZATION OF THE CATALYTIC ENZYME TO A PARTICULAR SUBCELLULAR  
COMPARTMENT.  
CC -I- SUBUNIT: PP2A CONSISTS OF A COMMON HETERODIMERIC CORE ENZYME,  
COMPOSED OF A 36 kDa CATALYTIC SUBUNIT (SUBUNIT C) AND A 65 kDa  
CONSTANT REGULATORY SUBUNIT (PR65 OR SUBUNIT A), THAT ASSOCIATES  
WITH A VARIETY OF REGULATORY SUBUNITS. PROTEINS THAT ASSOCIATE  
WITH THE CORE DIMER INCLUDE THREE FAMILIES OF REGULATORY SUBUNITS  
B (THE R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 AND R5/B'/B56  
FAMILIES), THE 48 kDa VARIABLE REGULATORY SUBUNIT, VIRAL PROTEINS,  
AND CELL SIGNALING MOLECULES.  
CC -I- TISSUE SPECIFICITY: BRAIN.  
CC -I- SIMILARITY: BELONGS TO THE PHOSPHATASE 2A REGULATORY SUBUNIT B  
FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Z34933; CAA84405.1;  
DR InterPro: IPR000009; PP2A\_PR55.  
DR Pfam: PF00400; WD40\_5.  
DR PRINTS: PRO0600; PP2APR55.  
DR SMART: SM00320; WD40\_4.  
DR PROSITE: PS01024; PR55\_1; 1.  
DR PROSITE: PS01025; PR55\_2; 1.  
KW Multigene family.  
SQ SEQUENCE 443 AA; 51459 MW; F8562FC696719F41 CRC64;

Query Match 73.5%; Score 25; DB 1; Length 443;  
Best Local Similarity 62.5%; Pred. No. 86;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVMXVAF 8  
||: |||  
DB 223 EYTAAEF 230

Search completed: October 30, 2002, 12:27:52  
Job time : 2.68796 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:20:50 ; Search time 1.60197 seconds

(without alignments)  
277.344 Million cell updates/sec

Title: US-09-724-571-103

Perfect score: 20

Sequence: 1 VKMD 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_032802:\*

```
1: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	21	AA832122
2	20	100.0	4	21	AA87945
3	20	100.0	5	18	AA808216
4	20	100.0	5	19	AA861150
5	20	100.0	5	20	AA833750
6	20	100.0	5	22	AA847260
7	20	100.0	8	14	AA842401
8	20	100.0	8	21	AA894772
9	20	100.0	8	22	AA810660
10	20	100.0	8	22	AA806902
11	20	100.0	8	22	AA806631

12	20	100.0	8	22	AA806635	Synthetic fluoresce
13	20	100.0	8	22	AA807230	Human beta-amyloid
14	20	100.0	8	22	AA802612	Human Aspartyl pro
15	20	100.0	9	19	AA882084	Fluorogenic protea
16	20	100.0	9	19	AA882083	Fluorogenic protea
17	20	100.0	9	21	AA807873	A peptide fragment
18	20	100.0	9	21	AA87949	Mammalian amyloid
19	20	100.0	10	13	AA822054	Peptide p1. Synth
20	20	100.0	10	13	AA824261	Human amyloid pr
21	20	100.0	10	20	AA882440	Human amyloid beta
22	20	100.0	10	21	AA867703	Beta-Ap alpha-sec
23	20	100.0	10	22	AA810654	Human wild-type AP
24	20	100.0	10	22	AA806899	Human amyloid prec
25	20	100.0	10	22	AA806828	ASP2 recognition s
26	20	100.0	10	22	AA807227	Human beta-amyloid
27	20	100.0	10	22	AA862668	Beta-sheet breaker
28	20	100.0	10	22	AA802606	Human wild-type AP
29	20	100.0	10	22	AA866574	Synthetic peptide
30	20	100.0	10	22	AA846205	Human APP derived
31	20	100.0	10	22	AA846206	Human APP derived
32	20	100.0	10	22	AA846207	Human APP derived
33	20	100.0	10	22	AA846208	Human APP derived
34	20	100.0	10	22	AA846209	Human APP derived
35	20	100.0	10	22	AA846210	Human APP derived
36	20	100.0	10	22	AA861336	Synthetic peptide f
37	20	100.0	11	22	AA875143	ASP 1 substrate se
38	20	100.0	11	22	AA875144	ASP 1 substrate se
39	20	100.0	11	22	AA874468	Beta-amyloid precu
40	20	100.0	12	22	AA874931	Beta-amyloid precu
41	20	100.0	13	19	AA870869	Quenched fluoresce
42	20	100.0	15	22	AA810669	Human Asp-2b activ
43	20	100.0	15	22	AA806906	Human beta-amyloid
44	20	100.0	15	22	AA807234	Human beta-amyloid
45	20	100.0	16	21	AA806315	Human beta-amyloid

## ALIGNMENTS

## RESULT 1

AA832122 standard; peptide: 4 AA.

AA832122;

14-FEB-2001 (first entry)

Beta-secretase target region of amyloid-beta precursor protein.

Aldehyde caspase inhibitor; antibody; amyloid-beta precursor;

caspase cleavage; neurodegenerative disease.

Synthetic.

WO200063250-A1.

26-OCT-2000.

13-APR-2000; 2000WO-CA00414.

15-APR-1999; 99US-0129495.

(MERI ) MERCK PROSST CANADA & CO.

Gervais F, Roy S, Nicholson DW, Xu D, Robertson G, Huang J;

WPI: 2000-687160/67.

Novel antibody that recognizes neo-epitopes of caspase cleaved  
amyloid-beta precursor protein useful for diagnostic conditions  
involving neuronal apoptosis such as Alzheimer's, Huntington's and  
Parkinson's disease -

PS Disclosure; Page 7; 65pp; English.

XX CC The present invention describes an antibody which recognises a  
CC neo-epitope created by the caspase mediated cleavage of amyloid-beta  
CC precursor protein (APP) or amyloid-beta precursor like protein (APLP).  
CC These proteins are involved in the premature death of cells at sites of  
CC neurodegeneration. The antibody of the invention can thus be used in the  
CC treatment of neurodegenerative diseases such as Alzheimer's, Huntington's  
CC and Parkinson's diseases, amyotrophic lateral sclerosis, progressive  
CC multiple sclerosis, head trauma, prion related conditions,  
CC Creutzfeldt-Jacob disease, spongiform encephalopathy, Friedreich's ataxia,  
CC fatal familial insomnia, Pelizaeus-Merzbacher disease, schizophrenia,  
CC dentatorubropallidolysian atrophy, spinocerebellar atrophy type 3,  
CC spinal bulbar muscular atrophy, spinal cord injury, stroke and brain  
CC injury.

XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 21; Length 4;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
|||  
DB 1 VKMD 4

RESULT 2  
AAV87945  
ID AAV87945 standard; protein; 4 AA.  
XX  
XX AAV87945;  
AC  
XX  
XX 11-SEP-2000 (first entry)  
DT

XX Mammalian APP Beta-secretase substrate peptide.  
DE  
XX Amyloid precursor protein; APP; secretase; vesicle; Abeta peptide;  
KM  
XX Alzheimer's disease.  
KW  
XX Mammalia.  
OS  
XX WO200023576-A2.  
PN  
XX 27-APR-2000.  
PD  
XX 15-OCT-1999; 99WO-US24403.  
PF  
XX 16-OCT-1998; 98US-0173887.  
PR  
XX 20-APR-1999; 99US-0294987.  
PS  
XX (HOOK/) HOOK V Y H.  
PA  
XX  
XX Hook VYH;  
PI  
XX  
XX WPI; 2000-339679/29.  
DR  
XX  
XX Determining the proteolytic activity of secretase for treating  
PT Alzheimer's disease comprises permeabilizing vesicles and incubating  
PT with amyloid precursor protein (APP) to determine cleavage of APP  
PT substrate -  
PS  
XX  
XX Disclosure; Page 96; 97pp; English.

XX This invention describes a novel method for the determination of  
CC the proteolytic activity of a secretase comprising obtaining and  
CC permeabilizing pure vesicles, incubating the vesicles with an amyloid  
CC precursor protein (APP) and determining the cleavage of the APP  
CC substrate where the amount of cleavage is proportional to the  
CC proteolytic activity of the secretase. The methods are useful for  
CC selecting secretases and agents that cleave the amyloid precursor  
CC protein substrate, inhibiting production of the Abeta peptide found  
CC in Alzheimer's disease and treating Alzheimer's disease in patients.

CC This sequence represents a fragment of mammalian amyloid precursor  
CC protein, APP, beta-secretase substrate which is used in the method  
CC of the invention.

XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 21; Length 4;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
|||  
DB 1 VKMD 4

RESULT 3  
AAW08216  
ID AAW08216 standard; peptide; 5 AA.  
XX  
XX AAW08216;  
AC  
XX  
XX 05-SEP-1997 (first entry)  
DT

XX Wild type APP beta-cleavage site #1.  
DE  
XX Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;  
KM alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.  
KW  
XX Homo sapiens.  
OS  
XX WO9640885-A2.  
PN  
XX 19-DEC-1996.  
PD  
XX 07-JUN-1996; 96WO-US09985.  
PF  
XX 07-JUN-1995; 95US-0485152.  
PR  
XX 07-JUN-1995; 95US-0480498.  
PS  
XX (ATHE-) ATHENA NEUROSCIENCES INC.  
PA  
XX  
XX Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;  
PI Mcconlogue LC, Sinha S, Tan H;  
PI  
XX  
XX WPI; 1997-052304/05.  
DR  
XX  
XX Beta-secretase which specifically cleaves beta-amyloid precursor  
PT protein - useful to screen for inhibitors useful in treatment of  
PT Alzheimer's disease  
PT  
XX  
XX Claim 5; Page 59; 92pp; English.

XX AAW08216, AAW08217 and AAW08350 represent beta-cleavage sites from  
CC beta-amyloid precursor proteins (APP). These sequences are recognised by  
CC the enzyme of the invention. The enzyme of the invention is  
CC beta-secretase, and specifically cleaves beta-APP at one of these sites.  
CC Normal processing of beta-APP is thought to occur via cleavage between  
CC residues 16 and 17 of the beta-amyloid peptide region by an  
CC alpha-secretase. Pathogenic processing is thought to occur by  
CC beta-secretase cleavage of beta-APP. Beta-secretase activity can be  
CC detected and measured using a method of the invention, which detects at  
CC least one of the beta-secretase cleavage products formed on cleavage. The  
CC method can be used to determine whether a test substance inhibits  
CC proteolytic cleavage, by beta-secretase, of beta-APP. Compounds effective  
CC to at least partially inhibit beta-secretase activity can be used to  
CC inhibit cleavage of beta-APP in cells or mammalian hosts. Isolation and  
CC purification of beta-secretase will permit chemical modelling of a  
CC critical event in the pathology of Alzheimer's disease.

XX SQ Sequence 5 AA;

Query Match 100.0%; Score 20; DB 18; Length 5;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;



```

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 VKMD 4
   ||||
Db 1 VKMD 4

RESULT 4
AAM61150
ID AAM61150 standard; Peptide: 5 AA.
AC AAM61150;
XX
XX 26-OCT-1998 (first entry)
XX
XX Wild-type APP cleavage site.
XX
XX Beta-secretase; human; beta-amyloid precursor protein; APP;
XX protease; inhibitor; screening; Alzheimer's disease; therapy.
XX
XX Homo sapiens.
XX
XX WO9826059-A1.
XX
XX 18-JUN-1998.
XX
XX 11-DEC-1996; 96WO-US19549.
XX
XX 11-DEC-1996; 96WO-US19549.
XX
XX (ATHE-) ATHENA NEUROSCIENCES INC.
XX
XX Anderson JP, Chrysler SMS, Kelm PS, Sinha S;
XX
XX WPI: 1998-348519/30.
XX
XX Novel beta-secretase which cleaves beta-amyloid precursor protein -
XX useful for screening for compounds which inhibit the cleavage and
XX are useful for treating Alzheimer's disease
XX
XX Disclosure: Page 20; 39pp; English.
XX
XX This peptide comprises the site of wild-type beta-amyloid precursor
XX protein (APP) (MBP-C125 WT) that is cleaved by a novel
XX beta-secretase isolated from human 293 cells. This protease
XX cleaves APP at the N-terminus of the beta-amyloid peptide (beta-AP).
XX It is believed to be the putative beta-secretase responsible for
XX the pathogenic processing of APP to beta-AP in Alzheimer's disease,
XX Down's syndrome and HCHWA-D. Recombinant fusion proteins were
XX generated comprising the last 125 amino acids of APP (wild-type
XX or Swedish double mutation, see AAM61151-52) fused to the C-terminal
XX end of maltose binding protein. The 2 fusion proteins were
XX expressed in Escherichia coli, and used as substrates for
XX beta-secretase in beta-secretase inhibitor assays. Compounds that
XX inhibit APP cleavage by beta-secretase may be useful in the
XX treatment of Alzheimer's disease.
XX
XX Sequence 5 AA:
XX
XX Query Match 100.0%; Score 20; DB 19; Length 5;
XX Best Local Similarity 100.0%; Pred. No. 6.4e+05;
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 VKMD 4
   ||||
Db 1 VKMD 4

RESULT 5
AAY33750
ID AAY33750 standard; Protein: 5 AA.
XX
XX AAY33750;
AC

```

```

XX
XX 09-NOV-1999 (first entry)
XX
XX Wild type beta-amyloid protein precursor (APP) cleavage site.
XX
XX Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
XX Alzheimer's disease; cleavage site; wild type.
XX
XX Homo sapiens.
XX
XX US5942400-A.
XX
XX 24-AUG-1999.
XX
XX 07-JUN-1996; 96US-0659984.
XX
XX 07-JUN-1996; 96US-0659984.
XX
XX 07-JUN-1995; 95US-0480498.
XX
XX 07-JUN-1995; 95US-0485152.
XX
XX (ELAN-) ELAN PHARM INC.
XX
XX Anderson JP, Jacobson-Croak KL, Sinha S;
XX
XX WPI: 1999-517417/43.
XX
XX A method for detecting human beta-secretase cleavage of polypeptides
XX useful for identifying beta-secretase inhibitors
XX
XX Examples: Column 28; 43pp; English.
XX
XX This sequence is the wild type beta-amyloid protein precursor (APP)
XX cleavage site. APP is cleaved by beta-secretase AAY33741. The wild type
XX cleavage site and the Swedish mutant version of this site AAY33751 are
XX used in a method for detecting human beta-secretase cleavage of
XX polypeptides and for identifying beta-secretase inhibitors. Inhibition
XX of beta-secretase activity would be useful for chemical modelling of a
XX critical event in the pathology of Alzheimer's disease. Inhibitors of
XX beta-secretase would be useful for the prevention and treatment of
XX Alzheimer's disease and Down's Syndrome.
XX
XX Sequence 5 AA:
XX
XX Query Match 100.0%; Score 20; DB 20; Length 5;
XX Best Local Similarity 100.0%; Pred. No. 6.4e+05;
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 VKMD 4
   ||||
Db 1 VKMD 4

RESULT 6
AAB47260
ID AAB47260 standard; Peptide: 5 AA.
AC AAB47260;
XX
XX 18-JUL-2001 (first entry)
XX
XX Wild type APP sequence for cleavage by beta-secretase.
XX
XX Beta-secretase; isotype; beta-amyloid precursor protein; APP;
XX beta-amyloid peptide; beta-AP; Alzheimer's disease; Down's syndrome;
XX HCHWA-D; Swedish mutation; maltose binding protein; MBP.
XX
XX Homo sapiens.
XX
XX US6221645-B1.
XX
XX 24-APR-2001.
XX
XX 07-JUN-1996; 96US-0660531.
XX
XX

```

```

XX 07-JUN-1995; 95US-0480498.
PR (ELAN-) ELAN PHARM INC.
XX Chryslers SMS, Sinha S, Keim PS, Anderson JP, Tan H, McConlogue LC;
XX WPI: 2001-315578/33.
DR Novel antibody that specifically binds native beta-secretase protein,
XX useful for raising anti-idiotypic antibodies and for detecting or
XX diagnosing pathological conditions related to presence of respective
XX antigens
XX Example; Column 28; 42pp; English.
XX The sequences given in AAB47260-61 represent cleavage sites derived
XX from wild-type and the Swedish mutation of beta-amyloid precursor
XX protein (APP). These cleavage sites were used in fusion proteins
XX which were used as substrates for the beta-secretase protein which
XX is characterized by an ability to cleave the 695-amino acid isoform
XX of APP between amino acids 596 and 597. The fusion proteins contain
XX the carboxy-terminal end of Maltose binding protein (MBP) fused to
XX the carboxy-terminal 125 amino acids of either wild type APP or APP
XX containing the Swedish mutation. Beta-secretase is thought to be
XX responsible for the pathogenic processing of APP to form beta amyloid
XX peptide (beta-AP) in beta-AP related conditions, e.g. Alzheimer's
XX disease, Down's syndrome, HCHWA-D etc. Beta-secretase has a molecular
XX weight of 260-300 kD and will bind to wheat germ agglutinin but not to
XX concanavalin A. Beta-secretase will cleave both the wild type and
XX the Swedish mutation of APP.
XX Sequence 5 AA:
SQ
Query Match 100.0%; Score 20; DB 22; Length 5;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VKMD 4
DB 1 VKMD 4
RESULT 7
AAB42401
ID AAB42401 standard; peptide; 8 AA.
XX AAB42401;
AC 01-APR-1994 (first entry)
XX 01-APR-1994 (first entry)
XX Peptide representing a variation of the C-terminus of ATF-beta APP.
XX Amyloid precursor protein; Alzheimer's disease; fragments; inhibitors;
XX amino-terminal fragment; diagnosis; screening; chemiluminescence assay.
XX Synthetic.
XX WO9321526-A.
XX 28-OCT-1993.
XX 03-MAR-1993; 93WO-US01817.
XX 15-APR-1992; 92US-0868949.
XX 26-OCT-1992; 92US-0965971.
XX (ATHE-) ATHENA NEUROSCIENCES INC.
XX (ELIT) LILLY & CO ELI.
XX Fritz LC, Schenk DB, Seubert PA;
XX WPI: 1993-351873/44.

```

```

XX Monitoring beta amyloid precursor protein processing - involves
XX detecting soluble fragments from cleavage at amino terminals of
XX peptide, used to study Alzheimer's disease and potential drugs
XX for it
XX Disclosure; Page 7; 38pp; English.
XX The peptide represents a variant of the C-terminus of the beta-
XX amyloid precursor protein amino terminal fragment. Such a peptide
XX was used in conjunction with antibody 92 in a chemiluminescence assay
XX to probe human lumbar cerebrospinal fluid to detect Alzheimer's
XX disease. Such an assay can be used to diagnose or monitor amyloid-
XX related diseases in a patient and to screen and evaluate potential
XX drugs for the treatment of these diseases.
XX See also AAB42398-404.
XX Sequence 8 AA:
SQ
Query Match 100.0%; Score 20; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VKMD 4
DB 5 VKMD 8
RESULT 8
AAY94772
ID AAY94772 standard; Protein; 8 AA.
XX AAY94772;
AC 12-FEB-2001 (first entry)
XX 12-FEB-2001 (first entry)
XX Beta-secretase substrate peptide SEQ ID 18.
XX Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease;
XX Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.
XX Synthetic.
XX WO200058479-A1.
XX 05-OCT-2000.
XX 23-MAR-2000; 2000WO-US07755.
XX 26-MAR-1999; 99US-0277229.
XX (AMGE-) AMGEN INC.
XX Citron M, Vassar RJ, Bennett BD;
XX WPI: 2000-594643/56.
XX Isolated beta-secretase nucleic acids and encoded polypeptides, useful
XX for diagnosis and gene therapy of Alzheimer's disease -
XX Example 10; Page 117; 145pp; English.
XX This invention relates to 3 nucleotide sequences encoding beta-secretase
XX proteins. Beta-secretase is an enzyme involved in the production of one
XX of the components of amyloid plaques involved in Alzheimer's disease. The
XX invention includes an expression vector comprising the nucleotide
XX sequence, a host cell comprising the expression vector, and a process for
XX producing the protein through culturing the transformed cells. Also
XX included in the invention are a polypeptide derivative of the
XX beta-secretase protein, a fusion protein comprising beta-secretase fused
XX to a heterologous amino acid sequence, and a method for modulating the
XX levels of beta-secretase polypeptide in a mammal comprising administering
XX the polynucleotide sequence. Beta-secretase exhibits neuroprotective and

```

CC nootropic activity. The beta-secretase nucleotide sequence may be used to  
CC map locations of the beta-secretase gene and related genes on chromosomes  
CC and as hybridization probes in diagnostic assays to test for the presence  
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's  
CC syndrome, and amyloid sequence. The nucleotide sequence may also be  
CC used as anti-sense inhibitors of beta-secretase expression, in gene  
CC therapy of Alzheimer's disease, and for the identification of compounds  
CC that modulate beta-secretase activity. Antibodies to the beta-secretase  
CC protein may be used for in vitro and in vivo diagnostic purposes to  
CC detect the presence of beta-secretase polypeptide in a body fluid or cell  
CC sample. The present sequence represents a beta-secretase substrate  
CC peptide.  
CC  
XX  
SQ Sequence 8 AA;  
  
Query Match 100.0%; Score 20; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VKMD 4  
| | | |  
Db 2 VKMD 5  
  
RESULT 9  
AAE10660  
ID AAE10660 standard; peptide: 8 AA.  
XX  
AC AAE10660;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE Human Aspartyl protease-1 (hu-Asp-1) beta-secretase, wild-type peptide.  
XX  
KW Human; aspartyl protease 1; Aspl; amyloid precursor protein; APP;  
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;  
KW amyloid plaque; neuronal loss; proteolytic; nootropic; neuroprotective.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Cleavage-site 4.5  
XX  
PN GB357767-A.  
XX  
PD 04-JUL-2001.  
XX  
PE 22-SEP-2000; 2000GB-0023315.  
XX  
PR 23-SEP-1999; 99US-0155493.  
PR 23-SEP-1999; 99US-0404133.  
PR 23-SEP-1999; 99MO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Bienkowski MJ, Gurney M;  
XX  
DR WPI: 2001-444208/48.  
XX  
PT Polypeptide comprising fragments of human aspartyl protease with  
PT amyloid precursor protein processing activity and alpha-secretase  
PT activity, for identifying modulators useful in treating Alzheimer's  
PT disease -  
XX  
PS Example 15; Page 92; 187pp; English.  
XX  
CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified  
CC Aspl proteins which lack transmembrane domain or amino terminal  
CC domain or cytoplasmic domain and retains alpha-secretase activity  
CC and amyloid protein precursor (APP) processing activity. The proteins  
CC of the invention are useful for assaying hu-Aspl alpha-secretase

CC activity, which in turn is useful for identifying modulators of  
CC hu-Aspl alpha-secretase activity, where modulators that increase  
CC hu-Aspl alpha-secretase activity are useful for treating Alzheimer's  
CC disease (AD) which causes progressive dementia with consequent  
CC formation of amyloid plaques, neurofibrillary tangles, gliosis and  
CC neuronal loss. Hu-Aspl protease substrate is useful for assaying  
CC hu-Aspl proteolytic activity, by contacting hu-Aspl protein with  
CC the substrate under acidic conditions and determining the level of  
CC hu-Aspl proteolytic activity. The present sequence is human aspartyl  
CC protease-1 (hu-Asp-1) beta-secretase, wild-type peptide which is used  
CC for determining the enzymatic activity of Aspl protein lacking a  
CC transmembrane (TM) domain and containing (His)6 tag.  
CC  
XX  
SQ Sequence 8 AA;  
  
Query Match 100.0%; Score 20; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VKMD 4  
| | | |  
Db 2 VKMD 5  
  
RESULT 10  
AAE06902  
ID AAE06902 standard; peptide: 8 AA.  
XX  
AC AAE06902;  
XX  
DT 23-OCT-2001 (first entry)  
XX  
DE Human amyloid precursor protein (APP) substrate peptide.  
XX  
KW Human; aspartyl protease 2; Asp 2; beta-amyloid precursor protein; APP;  
KW beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;  
KW neurofibrillary tangle; neuronal loss; amyloid-beta peptide; nootropic;  
KW neuroprotective; antisense therapy; gene therapy.  
OS Homo sapiens.  
XX  
FH WO200150829-A2.  
XX  
PN 19-JUL-2001.  
XX  
PD 09-MAY-2001; 2001WO-IB00799.  
XX  
PE 09-MAY-2001; 2001WO-IB00799.  
XX  
PR 09-MAY-2001; 2001WO-IB00799.  
XX  
PA (BIEN/) BIENKOWSKI M J.  
PA (GURN/) GURNEY M E.  
PA (HEIN/) HEINRIKSON R L.  
PA (PARO/) PARODI L A.  
PA (YANR/) YAN R.  
XX  
PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
XX  
DR WPI: 2001-483072/52.  
XX  
PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity -  
XX  
PS Claim 128; Page 101; 185pp; English.  
XX  
CC The invention relates to human aspartyl proteases (Hu-Asp), beta-amyloid  
CC precursor protein (APP) isoforms and their corresponding DNA molecules.  
CC Human aspartyl proteases can act as beta-secretase proteases useful for  
CC treating Alzheimer's disease. APP isoforms are useful for identifying  
CC modulators of amyloid-beta peptide production, for use in designing  
CC therapeutics for the treatment and prevention of Alzheimer's disease,  
CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis

CC and neuronal loss. APP isoforms are also used in methods for identifying  
 CC inhibitors and modulators of human Asp2 activity. The invention relates  
 CC to a method for identifying agents that modulate the activity of human  
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
 CC as a means to screen in cellular assays for the inhibitors of beta- and  
 CC gamma-secretase. Hu-Asp DNA fragments are useful as probes or primers in  
 CC polymerase chain reactions (PCR). The probes are useful for detecting  
 CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
 CC blots. The present sequence is human amyloid precursor protein (APP)  
 CC substrate peptide related to the invention.

XX Sequence 8 AA;

Query Match Best Local Similarity 100.0%; Score 20; DB 22; Length 8;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 Db 2 VKMD 5

RESULT 11  
 AAU06631  
 ID AAU06631 standard; Peptide: 8 AA.

AC AAU06631;

DT 24-OCT-2001 (first entry)

DE Beta secretase substrate peptide.

XX Aspartyl protease; Asp2; beta-secretase; nototropic;

KM neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;

KW amyloid-beta; Abeta; Beta secretase substrate peptide.

OS Synthetic.

PN WO200149098-A2.

PD 12-JUL-2001.

PF 09-MAY-2001; 2001WO-IB00798.

PR 09-MAY-2001; 2001WO-IB00798.

PA (BIEN/) BIENKOWSKI M J.

PA (GURN/) GURNEY M E.

PA (HEIN/) HEINRIKSON R L.

PA (PARO/) PARODI L A.

PA (YANR/) YAN R.

PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;

DR WPI; 2001-502549/55.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl

PT protease 2, lacking Asp2 transmembrane domain and retaining beta

PT secretase activity of Asp2 useful for identifying inhibitors of Asp2

PT activity

XX Claim 88; Page 94; 185pp; English.

XX The invention relates to a purified polypeptide comprising a fragment of

CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2

CC transmembrane domain and the Asp2 protein, and where the polypeptide and

CC the fragment retain the beta-secretase activity of the mammalian Asp2

CC protein. The invention also details polynucleotides for the Asp

CC proteins and vectors expressing them, and a polypeptide (isoform of

CC amyloid protein precursor (APP) comprising the amino acid sequence of an

CC APP or its fragment containing an APP cleavage site recognizable by a

CC mammalian beta-secretase, and further comprising two lysine residues at

CC the carboxyl terminus of the amino acid sequence of the mammalian APP or

CC APP fragment. Also included in the invention are methods of identifying  
 CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
 CC useful for treating Alzheimer's disease. APP is useful in methods for  
 CC identifying inhibitors or modulators of human Asp2 activity and  
 CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
 CC therapeutics for the treatment or prevention of Alzheimer's disease.  
 CC APP comprising the APP-SW-beta-secretase peptide sequence (NLDA), which  
 CC is associated with increased levels of Abeta processing is useful in  
 CC assays relating the Alzheimer's research. The expression vector is useful  
 CC for recombinantly expressing APP. Nucleic acids that hybridize to  
 CC Asp oligonucleotides are useful as probes or primers. The probes are  
 CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
 CC Northern and Southern blots. The present sequence is a beta secretase  
 CC substrate peptide.

XX Sequence 8 AA;

Query Match Best Local Similarity 100.0%; Score 20; DB 22; Length 8;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 Db 2 VKMD 5

RESULT 12  
 AAU06635  
 ID AAU06635 standard; Peptide: 8 AA.

AC AAU06635;

DT 24-OCT-2001 (first entry)

DE Synthetic fluorescent Asp2 substrate.

XX Aspartyl protease; Asp2; beta-secretase; nototropic;

KM neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;

KW amyloid-beta; Abeta.

OS Synthetic.

PN WO200149098-A2.

PD 12-JUL-2001.

PF 09-MAY-2001; 2001WO-IB00798.

PR 09-MAY-2001; 2001WO-IB00798.

PA (BIEN/) BIENKOWSKI M J.

PA (GURN/) GURNEY M E.

PA (HEIN/) HEINRIKSON R L.

PA (PARO/) PARODI L A.

PA (YANR/) YAN R.

PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;

DR WPI; 2001-502549/55.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl

PT protease 2, lacking Asp2 transmembrane domain and retaining beta

PT secretase activity of Asp2 useful for identifying inhibitors of Asp2

PT activity

XX Location/Qualifiers

FT Modified-site 1

FT /note="Glu is covalently linked to a fluorescent

FT MCA moiety"

FT Modified-site 8

FT /note="Glu is covalently linked to a fluorescent

FT K-DNP moiety"

PS Example 12; Page 81; 185pp; English.  
XX  
CC The invention relates to a purified polypeptide comprising a fragment of  
CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2  
CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
CC the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. The invention also details polynucleotides for the Asp  
CC proteins and vectors expressing them, and a polypeptide (isoform of  
CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
CC APP or its fragment containing an APP cleavage site recognizable by a  
CC mammalian beta-secretase, and further comprising two lysine residues at  
CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
CC APP fragment. Also included in the invention are methods of identifying  
CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
CC useful for treating Alzheimer's disease. APP is useful in methods for  
CC identifying inhibitors or modulators of human Asp2 activity and  
CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
CC therapeutics for the treatment or prevention of Alzheimer's disease.  
CC APP comprising the APP-Su-beta-secretase peptide sequence (NDA), which  
CC is associated with increased levels of Abeta processing is useful in  
CC assays relating to the Alzheimer's research. The expression vector is useful  
CC for recombinantly expressing APP. Nucleic acids that hybridize to  
CC APP oligonucleotides are useful as probes or primers. The probes are  
CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence is a synthetic  
CC fluorescent substrate used to assay Asp2.  
XX  
SQ Sequence 8 AA:  
Query Match 100.0%; Score 20; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VKMD 4  
   ||||  
   2 VKMD 5  
Db  
RESULT 13  
AAU07230  
ID AAU07230 standard; Peptide: 8 AA.  
XX  
AC AAU07230;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human beta-amyloid protein precursor; APP-beta secretase site peptide #3.  
XX  
KW Human; aspartyl protease 1; Asp-1; neurotrophic; neuroprotective;  
KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;  
KW beta-secretase; Alzheimer's disease; App-beta.  
XX  
OS Homo sapiens.  
XX  
PN WO200149097-A2.  
XX  
PD 12-JUL-2001.  
XX  
PF 09-MAY-2001; 2001WO-IB00797.  
XX  
PR 09-MAY-2001; 2001WO-IB00797.  
XX  
PA (BIEN/) BIENKOWSKI M J.  
PA (GURN/) GURNEY M E.  
PA (HEIN/) HEINRIKSON R L.  
PA (PARO/) PARODI L A.  
PA (YANR/) YAN R.  
XX  
PI Blenkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
XX WPI: 2001-502548/55.  
XX  
PT Novel purified polypeptide comprising fragment of mammalian aspartyl

PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity  
XX  
XX Claim 88; Page 94; 185pp; English.  
XX  
PS The invention relates to a novel purified polypeptide comprising a  
CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
CC and the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. Also included is an isoform of amyloid protein precursor (APP)  
CC comprising the amino acid sequence of a APP or its fragment containing  
CC an APP cleavage site recognizable by a mammalian beta-secretase, and  
CC further comprising two lysine residues at the carboxyl terminus of the  
CC amino acid sequence of the mammalian APP or APP fragment. The  
CC polypeptides are used for assaying for modulators of beta-secretase  
CC activity; identifying agents that inhibit the APP processing activity  
CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that  
CC modulate the activity of Asp2; and for reducing cellular production of  
CC amyloid beta (Abeta) from APP. Agents identified by the above methods  
CC are useful for treating Alzheimer's disease; and for identifying  
CC modulators of amyloid-beta (Abeta) peptide production, for use in  
CC designing therapeutics for the treatment or prevention of Alzheimer's  
CC disease. Probes and primers derived from APP nucleic acid sequences  
CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence represents the  
CC amino acid sequence of human amyloid protein precursor, APP-beta  
CC secretase site peptide substrate #3 used in assays of human Asp2 beta-  
CC secretase activity.  
XX  
SQ Sequence 8 AA:  
Query Match 100.0%; Score 20; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VKMD 4  
   ||||  
   2 VKMD 5  
Db  
RESULT 14  
AAE02612  
ID AAE02612 standard; peptide: 8 AA.  
XX  
AC AAE02612;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Human Aspartyl protease-1 (Hu-Asp-1) beta-secretase, wild-type peptide.  
XX  
KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;  
KW Alzheimer's disease; antialzheimer's; aspartyl protease 1; Aspl;  
KW beta-secretase.  
XX  
OS Homo sapiens.  
XX  
PN WO200123533-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 22-SEP-2000; 2000WO-US26080.  
XX  
PR 23-SEP-1999; 99US-0155493.  
PR 23-SEP-1999; 99WO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.

PI Gurney M, Bienkowski MJ;  
 XX  
 DR WPI; 2001-290516/30.  
 XX  
 PT Enzymes that cleave the alpha-secretase site of the amyloid precursor  
 XX protein, useful for the treatment of Alzheimer's disease -  
 XX  
 XX Example 15; Page 94; 189pp; English.  
 CC The present invention relates to enzymes for cleaving the alpha-  
 CC secretase site of the amyloid precursor protein (APP) and methods of  
 CC identifying those enzymes. The methods may be used to identify enzymes  
 CC that may be used to cleave the alpha-secretase cleavage site of the APP  
 CC protein. The enzymes may be used to treat or modulate the progress of  
 CC Alzheimer's disease. The present sequence is human Aspartyl protease-1  
 CC (hu-ASP-1) beta-secretase, wild-type peptide which is used for  
 CC determining the enzymatic activity of Asp-1 deltatm (His)6 protein.  
 CC  
 XX Sequence 8 AA:  
 SQ  
 Query Match 100.0%; Score 20; DB 22; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VKMD 4  
 IIII  
 DB 2 VKMD 5  
 DB

RESULT 15  
 AAW82084  
 ID AAW82084 standard; peptide; 9 AA.  
 XX  
 AC AAW82084;  
 XX  
 DT 18-FEB-1999 (first entry)  
 XX  
 DE Fluorogenic protease indicator protease binding peptide #62.  
 XX  
 KW Protease activity; fluorophore; detection; fluorogenic; cellular uptake;  
 XX conformation change.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9837226-A1.  
 XX  
 PD 27-AUG-1998.  
 XX  
 PF 20-FEB-1998; 98WO-US03000.  
 XX  
 PR 20-FEB-1997; 97US-0802981.  
 XX  
 PA (ONCO-) ONCOIMMUNIN INC.  
 XX  
 PI Komoriya A, Packard BS;  
 XX  
 DR WPI; 1998-467579/40.  
 XX  
 PT New fluorogenic compositions - containing 2 fluorophores separated  
 PT by a peptide comprising a protease binding site, used for detecting  
 PT protease activity in samples.  
 XX  
 PS Claim 4; Page 77; 90pp; English.  
 XX  
 CC AAW82023-W82240 are peptides used in the construction of a fluorogenic  
 CC composition which is used for the detection of protease activity in  
 CC biological samples. The products can be used for the detection of  
 CC conformation changes in nucleic acids, oligosaccharides,  
 CC polysaccharides, proteins, lipids, phospholipids, glycolipids,  
 CC glycoproteins, steroids or polymers. In addition, attachment of a  
 CC hydrophobic group to a molecule can be used to enhance uptake by cells.  
 CC The composition is composed of P = peptide comprising a protease binding  
 CC site for the protease, F1, F2 peptides = fluorophores where F1 is

CC attached to the amino terminal amino acid and F2 is attached to the  
 CC carboxyl terminal amino acid and S1, S2 peptides = when present, are  
 CC peptide spacers where S1, when present, is attached to the amino terminal  
 CC acid, and S2, when present, is attached to the carboxyl terminal amino  
 CC acid.  
 XX  
 XX Sequence 9 AA:  
 SQ  
 Query Match 100.0%; Score 20; DB 19; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VKMD 4  
 IIII  
 DB 3 VKMD 6  
 DB

Search completed: October 30, 2002, 12:27:10  
 job time : 2.60197 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 ; Search time 0.570025 Seconds  
(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-103

Perfect score: 20

Sequence: 1 VKMD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	4	US-09-173-887-1	Sequence 1, Appli
2	20	100.0	4	US-09-294-987-2	Sequence 2, Appli
3	20	100.0	4	US-09-054-334-5	Sequence 5, Appli
4	20	100.0	5	US-08-480-498-1	Sequence 1, Appli
5	20	100.0	5	US-08-659-984A-13	Sequence 13, Appli
6	20	100.0	5	US-08-660-531-13	Sequence 13, Appli
7	20	100.0	5	US-09-054-334-1	Sequence 1, Appli
8	20	100.0	7	US-08-792-553-10	Sequence 10, Appli
9	20	100.0	8	US-07-965-971-4	Sequence 4, Appli
10	20	100.0	8	US-08-143-697-4	Sequence 4, Appli
11	20	100.0	8	US-08-440-261-4	Sequence 4, Appli
12	20	100.0	8	US-08-440-423-4	Sequence 4, Appli
13	20	100.0	8	US-08-846-444-4	Sequence 4, Appli
14	20	100.0	8	PCT-US94-07043A-4	Sequence 4, Appli
15	20	100.0	9	US-08-802-981-221	Sequence 221, App
16	20	100.0	9	US-08-802-981-222	Sequence 222, App
17	20	100.0	9	US-09-294-987-6	Sequence 6, Appli
18	20	100.0	10	US-08-025-321C-1	Sequence 1, Appli
19	20	100.0	11	PCT-US94-07043A-7	Sequence 7, Appli
20	20	100.0	12	PCT-US94-07043A-2	Sequence 2, Appli
21	20	100.0	16	PCT-US94-07043A-1	Sequence 1, Appli
22	20	100.0	18	US-08-149-975A-3	Sequence 3, Appli
23	20	100.0	21	US-08-802-981-114	Sequence 114, App
24	20	100.0	21	US-08-802-981-115	Sequence 115, App
25	20	100.0	27	US-08-141-324-11	Sequence 11, Appli
26	20	100.0	27	US-08-541-902-11	Sequence 11, Appli
27	20	100.0	27	US-08-936-165A-335	Sequence 335, App

28	20	100.0	45	1	US-08-462-859A-5	Sequence 5, Appli
29	20	100.0	45	1	US-08-123-659A-5	Sequence 5, Appli
30	20	100.0	45	1	US-08-464-247A-5	Sequence 5, Appli
31	20	100.0	45	1	US-08-464-248A-5	Sequence 5, Appli
32	20	100.0	53	4	US-09-173-887-5	Sequence 5, Appli
33	20	100.0	53	4	US-09-294-987-1	Sequence 1, Appli
34	20	100.0	58	1	US-08-371-930-25	Sequence 25, Appli
35	20	100.0	58	5	PCT-US94-01712-25	Sequence 25, Appli
36	20	100.0	59	1	US-08-484-969-3	Sequence 3, Appli
37	20	100.0	59	1	US-08-472-627-3	Sequence 3, Appli
38	20	100.0	59	1	US-08-388-463-3	Sequence 3, Appli
39	20	100.0	63	1	US-08-462-859A-3	Sequence 3, Appli
40	20	100.0	63	1	US-08-462-859A-4	Sequence 4, Appli
41	20	100.0	63	1	US-08-123-659A-3	Sequence 3, Appli
42	20	100.0	63	1	US-08-123-659A-4	Sequence 4, Appli
43	20	100.0	63	1	US-08-464-247A-3	Sequence 3, Appli
44	20	100.0	63	1	US-08-464-247A-4	Sequence 4, Appli
45	20	100.0	63	1	US-08-464-248A-3	Sequence 3, Appli

## ALIGNMENTS

RESULT 1  
US-09-173-887-1  
; Sequence 1, Application US/09173887  
; Patent No. 6245884  
; GENERAL INFORMATION:  
; APPLICANT: Hook, Vivian Y.H.  
; TITLE OF INVENTION: SECRETASES RELATED TO ALZHEIMER'S DEMENTIA  
; FILE REFERENCE: P-AS 3337  
; CURRENT APPLICATION NUMBER: US/09/173,887  
; CURRENT FILING DATE: 1998-10-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: mammalian  
US-09-173-887-1

Query Match 100.0%; Score 20; DB 4; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
DB 1 VKMD 4

RESULT 2  
US-09-294-987-2  
; Sequence 2, Application US/09294987  
; Patent No. 6313268  
; GENERAL INFORMATION:  
; APPLICANT: Hook, Vivian Y.H.  
; TITLE OF INVENTION: SECRETASES RELATED TO ALZHEIMER'S DEMENTIA  
; FILE REFERENCE: P-AS 3515  
; CURRENT APPLICATION NUMBER: US/09/294,987  
; CURRENT FILING DATE: 1999-04-20  
; PRIOR APPLICATION NUMBER: US 09/173,887  
; PRIOR FILING DATE: 1998-10-16  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: mammalian  
US-09-294-987-2

Query Match 100.0%; Score 20; DB 4; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
Db 1 VKMD 4

RESULT 3  
US-09-054-334-5  
Sequence 5, Application US/09054334  
Patent No. 6329163  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
APPLICANT: Sinha, Sukanto  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,334  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-0028200S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-054-334-5  
Query Match 100.0%; Score 20; DB 4; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VKMD 4  
Db 1 VKMD 4  
RESULT 4  
US-08-480-498-1  
Sequence 1, Application US/08480498  
Patent No. 5744346  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: One Market Plaza, Stewart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,498  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-002200  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-480-498-1

Query Match 100.0%; Score 20; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VKMD 4  
Db 1 VKMD 4

RESULT 5  
US-08-659-984A-13  
Sequence 13, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541



REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-13

Query Match 100.0%; Score 20; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VKMD 4  
Db 1 VKMD 4

RESULT 6  
US-08-660-531-13  
Sequence 13, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-13

Query Match 100.0%; Score 20; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VKMD 4

Db 1 VKMD 4

RESULT 7  
US-09-054-334-1  
Sequence 1, Application US/09054334  
Patent No. 6329163  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
APPLICANT: Sinha, Sukanto  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,334  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-002820US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-054-334-1

Query Match 100.0%; Score 20; DB 4; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VKMD 4  
Db 1 VKMD 4

RESULT 8  
US-08-792-553-10  
Sequence 10, Application US/08792553  
Patent No. 5981200  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Helm, Roger  
TITLE OF INVENTION: Tandem Fluorescent Protein Constructs  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: California

COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/792,553  
FILING DATE: 31-JAN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lisa A. Halle, Ph.D.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07257/041001/UC 96-160-2  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-792-553-10

Query Match 100.0%; Score 20; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
||||  
DB 1 VKMD 4

RESULT 9  
US-07-965-971-4  
Sequence 4, Application US/07965971  
Patent No. 5441870  
GENERAL INFORMATION:  
APPLICANT: Seubert, Peter A.  
APPLICANT: Schenk, Dale B.  
APPLICANT: Fritz, Lawrence C.  
TITLE OF INVENTION: Methods and Compositions for Monitoring  
TITLE OF INVENTION: Cellular Processing of Beta-Amyloid Precursor Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/965,971  
FILING DATE: 26-OCT-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/868,949  
FILING DATE: 15-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-4-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-965-971-4

Query Match 100.0%; Score 20; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
||||  
DB 5 VKMD 8

RESULT 10  
US-08-143-697-4  
Sequence 4, Application US/08143697  
Patent No. 5604102  
GENERAL INFORMATION:  
APPLICANT: Seubert, Peter A.  
APPLICANT: Schenk, Dale B.  
APPLICANT: Fritz, Lawrence C.  
TITLE OF INVENTION: Methods of Screening for Beta-Amyloid  
TITLE OF INVENTION: Peptide Production Inhibitors  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/143,697  
FILING DATE: 27-OCT-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/965,971  
FILING DATE: 26-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/868,949  
FILING DATE: 15-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-4-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-143-697-4

Query Match 100.0%; Score 20; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
||||

Db 5 VKMD 8

```
RESULT 11
US-08-440-261-4
; Sequence 4, Application US/08440261
; Patent No. 5605811
; GENERAL INFORMATION:
; APPLICANT: Seubert, Peter A.
; APPLICANT: Schenk, Dale B.
; APPLICANT: Filtz, Lawrence C.
; TITLE OF INVENTION: Methods and Compositions for Monitoring
; TITLE OF INVENTION: Cellular Processing of Beta-Amyloid Precursor Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Stewart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,261
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,971
; FILING DATE: 26-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-4-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-440-261-4

Query Match 100.0%; Score 20; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
Db 5 VKMD 8

RESULT 12
US-08-440-423-4
; Sequence 4, Application US/08440423
; Patent No. 5721130
; GENERAL INFORMATION:
; APPLICANT: Seubert, Peter A.
; APPLICANT: Schenk, Dale B.
; APPLICANT: Filtz, Lawrence C.
; TITLE OF INVENTION: Methods and Compositions for Monitoring
; TITLE OF INVENTION: Cellular Processing of Beta-Amyloid Precursor Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Stewart Tower, Suite 2000
; CITY: San Francisco
```

```
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,423
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,971
; FILING DATE: 26-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15270-4-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-440-423-4
```

```
Query Match 100.0%; Score 20; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
Db 5 VKMD 8
```

```
RESULT 13
US-08-846-444-4
; Sequence 4, Application US/08846444
; Patent No. 6018024
; GENERAL INFORMATION:
; APPLICANT: Seubert, Peter A.
; APPLICANT: Schenk, Dale B.
; APPLICANT: Filtz, Lawrence C.
; TITLE OF INVENTION: Methods and Compositions for Monitoring
; TITLE OF INVENTION: Cellular Processing of Beta-Amyloid Precursor Protein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,444
; FILING DATE: No. 6018024 yet assigned
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/868,949
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,971
; FILING DATE: 26-OCT-1992
```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US93/01817  
FILING DATE: 03-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/440,423  
FILING DATE: 12-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Storella, John R.  
REGISTRATION NUMBER: 32,944  
REFERENCE/DOCKET NUMBER: 015270-00045005  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
STRANDEDNESS:  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-846-444-4

Query Match  
Best Local Similarity 100.0%; Score 20; DB 3; Length 8;  
Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
|||||  
DB 5 VKMD 8

RESULT 14  
PCT-US94-07043A-4  
Sequence 4, Application PC/TUS9407043A  
GENERAL INFORMATION:  
APPLICANT: Tamburini, Paul P.; Benz, G nter; H blich,  
APPLICANT: Dieter; Dreyer, Robert N.; Koenig, Gerhard  
TITLE OF INVENTION: CATHEPSIN D IS AN AMYLOIDGENIC  
TITLE OF INVENTION: PROTEASE IN ALZHEIMER S DISEASE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Miles Inc.  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06516  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: Sharp PC 4600  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07043A  
FILING DATE: June 21, 1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10889  
FILING DATE: November 12, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/995,660  
FILING DATE: December 16, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/880,914  
FILING DATE: May 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pamela A. Simonton  
REGISTRATION NUMBER: 31,060  
REFERENCE/DOCKET NUMBER: MTI 224.3  
TELEPHONE: (203) 937-2340  
TELEFAX: (203) 937-2795  
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US94-07043A-4

Query Match  
Best Local Similarity 100.0%; Score 20; DB 5; Length 8;  
Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
|||||  
DB 5 VKMD 8

RESULT 15  
US-08-802-981-221  
Sequence 221, Application US/08802981  
GENERAL INFORMATION:  
PATENT NO. 6037137  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-00030005  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 221:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-802-981-221

Query Match  
Best Local Similarity 100.0%; Score 20; DB 3; Length 9;  
Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
|||||  
DB 3 VKMD 6

Search completed: October 30, 2002, 12:32:36  
Job time : 1.57002 secs

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 50.29 Seconds

(without alignments)  
1568.616 Million cell updates/sec

Title: US-09-724-571-43

Perfect score: 2419  
Sequence: 1 ETDEPEPEPRGRSGFVEMVD.....CLRCRQHQHDDFADISILK 456

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacteriopl:\*  
17: SP-archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2419	100.0	532	4 Q9ULS1	Q9ULS1 homo sapien
2	2264.5	93.6	476	4 Q9BYC1	Q9BYC1 homo sapien
3	2161	89.3	457	4 Q9BYC0	Q9BYC0 homo sapien
4	2016.5	83.4	432	4 Q9BYB9	Q9BYB9 homo sapien
5	1407	58.2	266	11 Q9C0U5	Q9C0U5 mus musculu
6	1160.5	48.0	439	4 Q9J2V8	Q9J2V8 homo sapien
7	1146.5	47.4	514	11 Q9J1L8	Q9J1L8 mus musculu
8	972.5	40.2	468	4 Q9NZL2	Q9NZL2 homo sapien
9	967.5	40.0	396	4 Q9NZL1	Q9NZL1 homo sapien
10	712.5	29.5	213	4 Q9P0D2	Q9P0D2 homo sapien
11	596.5	24.7	255	11 Q9R1P7	Q9R1P7 mus musculu
12	332.5	13.7	391	5 Q9VKP6	Q9VKP6 drosophila
13	332	13.7	354	5 Q9GYX7	Q9GYX7 boophilus m
14	312.5	12.9	386	6 Q9BGC5	Q9BGC5 bos taurus
15	308	12.7	384	13 Q9DEC2	Q9DEC2 xenopus lae
16	308	12.7	385	13 Q9DEC4	Q9DEC4 rana catesb

17	305	12.6	386	6 Q9GMV7	Q9GMV7 rhinolophus
18	305	12.6	387	6 Q9GMV8	Q9GMV8 scroex ungui
19	304.5	12.6	372	5 Q9YLK3	Q9YLK3 drosophila
20	304.5	12.6	387	13 Q9DDV5	Q9DDV5 salvellinus
21	304	12.6	387	6 Q9GMV9	Q9GMV9 suncus murti
22	303.5	12.5	383	13 Q9DE45	Q9DE45 salvellinus
23	303	12.5	383	13 Q9DEC3	Q9DEC3 xenopus lae
24	302.5	12.5	376	13 Q9PUR8	Q9PUR8 pseudopleur
25	301	12.4	382	13 Q9PRG9	Q9PRG9 gallus galli
26	301	12.4	423	5 Q9YKP7	Q9YKP7 drosophila
27	298.5	12.3	384	13 Q913Z2	Q913Z2 rana catesb
28	295.5	12.2	386	6 Q9GMV6	Q9GMV6 canis famill
29	294	12.2	396	13 Q934Z8	Q934Z8 chionotraco
30	290.5	12.0	381	6 Q9GK11	Q9GK11 camelus dro
31	288	11.9	399	13 Q93458	Q93458 podarcis sl
32	287.5	11.9	444	5 Q21966	Q21966 caenorhabdi
33	284	11.7	398	13 P87370	P87370 oncorhynch
34	284	11.7	427	5 P91802	P91802 schistosoma
35	281	11.6	378	13 Q9PUR9	Q9PUR9 pseudopleur
36	279.5	11.6	390	6 Q9GK10	Q9GK10 camelus dro
37	278	11.5	370	6 Q9TWT1	Q9TWT1 bos taurus
38	278	11.5	399	13 Q9DD89	Q9DD89 brachydanio
39	278	11.5	422	5 Q96906	Q96906 onchocerca
40	277.5	11.5	380	6 Q28950	Q28950 sus scrofa
41	277	11.5	446	5 Q9N9H3	Q9N9H3 necator ame
42	276	11.4	389	13 Q9PMK1	Q9PMK1 gallus galli
43	276	11.4	389	13 Q9W643	Q9W643 gallus galli
44	275	11.4	396	13 Q9DEX3	Q9DEX3 clupea hare
45	275	11.4	385	6 Q29080	Q29080 sus scrofa

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	532 AA.
Q9ULS1	Q9ULS1	Q9ULS1		
AC	Q9ULS1	Q9ULS1		
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	KIA1149 PROTEIN (FRAGMENT).			
GN	KIA1149.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RX	MEDLINE=20039618; PubMed=10574461;			
RA	Hirosawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;			
RT	"Characterization of cDNA clones selected by the Genemart analysis			
RT	from size-fractionated cDNA libraries from human brain.";			
RL	DNA Res. 6:329-336(1999).			
DR	EMBL: AB032975; BAA6463.2; -			
DR	HSSP: P56272; IAW5.			
DR	MEROPS: A01.004; -			
DR	InterPro: IPR001969; Asp-protease.			
DR	InterPro: IPR001461; Pepsin.			
DR	Pfam: PF000026; asp. 1.			
DR	PRINTS: PR00792; PEPsin.			
DR	PROSITE: PS00141; ASP-PROTEASE; UNKNOWN_1.			
FT	NON TER			
SO	SEQUENCE	532 AA; 58720 MW; 96B135DD5FBD2E8 CRC64;		

Query Match 100.0%; Score 2419; DB 4; Length 532;  
Best Local Similarity 100.0%; Pred. No. 7, 3e-198;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ETDEPEPEPRGRSGFVEMVDNLKRGSGGYVEMVTGSPPTLNTLVDYTGSSNFAVGAP 60  
|||||

```

Db 77 ETDEPEEPGRGSRFVEMVDNLKRGSGQGYVEMTVGSPQTINLIIVDTGSSNFVAGAAP 136
QY 61 HPELHRYQROLSTYRDLRKGVYPTTQGWEGELGTDLYSIHPGVNVTYRANIAATE 120
Db 137 HPELHRYQROLSTYRDLRKGVYPTTQGWEGELGTDLYSIHPGVNVTYRANIAATE 196
QY 121 SDKPFINSNMEGILGLAYAEIARPDLSLEPFDSLVKOTHPNLFSLQCGAGPPLNOS 180
Db 197 SDKPFINSNMEGILGLAYAEIARPDLSLEPFDSLVKOTHPNLFSLQCGAGPPLNOS 256
QY 181 EYLASVSGSMITGGIDHSLYTGSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDK 240
Db 257 EYLASVSGSMITGGIDHSLYTGSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDK 316
QY 241 SYVDSGTTNLRPKKVFEEAAYKSIKAASSTKEKPDGFMGLGEOLVQWQAGTTPMNIFFPYS 300
Db 317 SYVDSGTTNLRPKKVFEEAAYKSIKAASSTKEKPDGFMGLGEOLVQWQAGTTPMNIFFPYS 376
QY 301 LYLMGEVTVNOSFRITILPQOYLRPVEDVATSDQDCYKFAISOSTGTVMGAVIMEGFYV 360
Db 377 LYLMGEVTVNOSFRITILPQOYLRPVEDVATSDQDCYKFAISOSTGTVMGAVIMEGFYV 436
QY 361 FDRARRKRIIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESLMTIAYVMAI 420
Db 437 FDRARRKRIIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESLMTIAYVMAI 496
QY 421 CALFMLPCLIAMVCOMRCLRCLROQHDHDFADISILK 456
Db 497 CALFMLPCLIAMVCOMRCLRCLROQHDHDFADISILK 532

RESULT 2
QYBYC1 PRELIMINARY: PRT: 476 AA.
AC Q9BYC1.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-476.
CN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production.";
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050436; BAB40931.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp--protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 476 AA; 52907 MW; 6C8C87E8A953AF66 CRC64;

Query Match 93.6%; Score 2264.5; DB 4; Length 476;
Best Local Similarity 94.5%; Pred. No. 9.5e-185;
Matches 431; Conservative 0; Mismatches 0; Indels 25; Gaps 1;
QY 1 ETDEPEEPGRGSRFVEMVDNLKRGSGQGYVEMTVGSPQTINLIIVDTGSSNFVAGAAP 60
Db 46 ETDEPEEPGRGSRFVEMVDNLKRGSGQGYVEMTVGSPQTINLIIVDTGSSNFVAGAAP 105
QY 61 HPELHRYQROLSTYRDLRKGVYPTTQGWEGELGTDLYSIHPGVNVTYRANIAATE 120

```

```

Db 106 HPELHRYQROLSTYRDLRKGVYPTTQGWEGELGTDLYSIHPGVNVTYRANIAATE 165
QY 121 SDKPFINSNMEGILGLAYAEIARPDLSLEPFDSLVKOTHPNLFSLQCGAGPPLNOS 180
Db 166 SDKPFINSNMEGILGLAYAEIARPDLSLEPFDSLVKOTHPNLFSLQCGAGPPLNOS 200
QY 181 EYLASVSGSMITGGIDHSLYTGSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDK 240
Db 201 EYLASVSGSMITGGIDHSLYTGSLWYTPIRREMYEVIIVVEINGODLKMCKEYNDK 260
QY 241 SYVDSGTTNLRPKKVFEEAAYKSIKAASSTKEKPDGFMGLGEOLVQWQAGTTPMNIFFPYS 300
Db 261 SYVDSGTTNLRPKKVFEEAAYKSIKAASSTKEKPDGFMGLGEOLVQWQAGTTPMNIFFPYS 320
QY 301 LYLMGEVTVNOSFRITILPQOYLRPVEDVATSDQDCYKFAISOSTGTVMGAVIMEGFYV 360
Db 321 LYLMGEVTVNOSFRITILPQOYLRPVEDVATSDQDCYKFAISOSTGTVMGAVIMEGFYV 380
QY 361 FDRARRKRIIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESLMTIAYVMAI 420
Db 381 FDRARRKRIIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESLMTIAYVMAI 440
QY 421 CALFMLPCLIAMVCOMRCLRCLROQHDHDFADISILK 456
Db 441 CALFMLPCLIAMVCOMRCLRCLROQHDHDFADISILK 476

RESULT 3
QYBYC0 PRELIMINARY: PRT: 457 AA.
AC Q9BYC0.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME
DE TYPE C).
CN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production.";
RL Neurosci. Lett. 307:9-12(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EXOCRINE PANCREAS;
RX Zaccchetti D., De Pietri Tonelli D., Schnurbus R.;
RT "New beta-site APP cleaving enzyme isoform (BACE-1c) obtained from
RT human pancreas.";
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB050437; BAB40932.1; -.
DR EMBL; AF338817; AAK38375.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp.protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 457 AA; 51068 MW; C794C9A9E8B5FE7A2 CRC64;

Query Match 89.3%; Score 2161; DB 4; Length 457;
Best Local Similarity 90.4%; Pred. No. 6.1e-176;
Matches 412; Conservative 0; Mismatches 0; Indels 44; Gaps 1;
QY 1 ETDEPEEPGRGSRFVEMVDNLKRGSGQGYVEMTVGSPQTINLIIVDTGSSNFVAGAAP 60

```

```

|||||
Db 46 ETDEPEEPERRRSFEKMDVNLNRKSGGGYVEMTVGSPQTILNIIVDGSSNFAGVGAAP 105
QY 61 HPELHRYRQRLSSTYRDLRKGYVVPYTGKMEGELGTDLVSIPHGPNVTVRANIAATE 120
Db 106 HPELHRYRQRLSSTYRDLRKGYVVPYTGKMEGELGTDL----- 145
QY 121 SDFEFGNSMWEGLGLAVAEIARPDLSLEPFDSLVKQTHVNLFSLOLCGAGFLNOS 180
Db 146 -----PDSLEPFDSLVKQTHVNLFSLOLCGAGFLNOS 181
QY 181 EYLASVGGSMITGGIDHSLYTGSMTPTIRREMYEVIIVREINQDLKMDCKEYNDK 240
Db 182 EYLASVGGSMITGGIDHSLYTGSMTPTIRREMYEVIIVREINQDLKMDCKEYNDK 241
QY 241 SIYDGTNTLRPKKVFEEAFAVKSIAASTEKFPDGFMLGEOLVQWQAGTTPMNIFFPVS 300
Db 242 SIYDGTNTLRPKKVFEEAFAVKSIAASTEKFPDGFMLGEOLVQWQAGTTPMNIFFPVS 301
QY 301 LYLMGEVTNOSFRITILPQOYLRPVEDVATSODDCYKFAISOSTGTGVGAVIMEGFYV 360
Db 302 LYLMGEVTNOSFRITILPQOYLRPVEDVATSODDCYKFAISOSTGTGVGAVIMEGFYV 361
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESTLMTIAYVMAI 420
Db 362 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESTLMTIAYVMAI 421
QY 421 CALFMLPLCLMWQWRCLRCLRQOHDFADDISLTK 456
Db 422 CALFMLPLCLMWQWRCLRCLRQOHDFADDISLTK 457

```

```

RESULT 4
Q9BYB9 PRELIMINARY; PRT; 432 AA.
ID Q9BYB9;
AC Q9BYB9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-432.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tadira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production."
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050438; BAB40933.1; -.
DR HSSP; P32329; 1YPS.
DR InterPro; IPR001969; Asp_protase.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp.1.
DR PRINTS; PR00792; Pepsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN.1.
SO SEQUENCE 432 AA; 48212 MW; 96FC81E6F0EED01B CRC64;

```

```

Query Match 83.4%; Score 2016.5; DB 4; Length 432;
Best Local Similarity 84.9%; Pred. No. 1.2e-163;
Matches 387; Conservative 0; Mismatches 0; Indels 69; Gaps 1;
QY 1 ETDEPEEPGRGSGFVEMVDNLNRKSGGGYVEMTVGSPQTILNIIVDGSSNFAGVGAAP 60
Db 46 ETDEPEEPGRGSGFVEMVDNLNRKSGGGYVEMTVGSPQTILNIIVDGSSNFAGVGAAP 105
QY 61 HPELHRYRQRLSSTYRDLRKGYVVPYTGKMEGELGTDLVSIPHGPNVTVRANIAATE 120

```

```

|||||
Db 106 HPELHRYRQRLSSTYRDLRKGYVVPYTGKMEGELGTDL----- 145
QY 121 SDFEFGNSMWEGLGLAVAEIARPDLSLEPFDSLVKQTHVNLFSLOLCGAGFLNOS 180
Db 146 -----LCGAGFLNOS 156
QY 181 EYLASVGGSMITGGIDHSLYTGSMTPTIRREMYEVIIVREINQDLKMDCKEYNDK 240
Db 157 EYLASVGGSMITGGIDHSLYTGSMTPTIRREMYEVIIVREINQDLKMDCKEYNDK 216
QY 241 SIYDGTNTLRPKKVFEEAFAVKSIAASTEKFPDGFMLGEOLVQWQAGTTPMNIFFPVS 300
Db 212 SIYDGTNTLRPKKVFEEAFAVKSIAASTEKFPDGFMLGEOLVQWQAGTTPMNIFFPVS 276
QY 301 LYLMGEVTNOSFRITILPQOYLRPVEDVATSODDCYKFAISOSTGTGVGAVIMEGFYV 360
Db 277 LYLMGEVTNOSFRITILPQOYLRPVEDVATSODDCYKFAISOSTGTGVGAVIMEGFYV 336
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESTLMTIAYVMAI 420
Db 337 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESTLMTIAYVMAI 396
QY 421 CALFMLPLCLMWQWRCLRCLRQOHDFADDISLTK 456
Db 397 CALFMLPLCLMWQWRCLRCLRQOHDFADDISLTK 432

```

```

RESULT 5
Q9CUC5 PRELIMINARY; PRT; 266 AA.
ID Q9CUC5;
AC Q9CUC5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME (FRAGMENT).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE-BRAIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Koichiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Mikaldo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderrelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazarrelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wyshnaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK014390; BAB2917.1; -.
DR MEROBS; A01.004; -.
DR MGD; MGI:1346342; Bace.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF000026; asp.2.
FT NON_TER 1
SO SEQUENCE 266 AA; 30201 MW; B913FDA8ADAB4238 CRC64;

```

Query Match 58.2%; Score 1407; DB 11; Length 266;  
 Best Local Similarity 98.9%; Pred. No. 5.9e-112;  
 Matches 263; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 191 IIGGIDHSLSYTGSLMYTPPIRREMYEVIIVVEINGDOLKMDCKEYNKDSIVSGTTL 250  
 DB 1 IIGGIDHSLSYTGSLMYTPPIRREMYEVIIVVEINGDOLKMDCKEYNKDSIVSGTTL 60

QY 251 RLPKRVFAAASKIKAASTEFKPPGCFMLGBOIYVMOAGTTPMNIFFPISLYLMEVTNQ 310  
 DB 61 RLPKRVFAAASKIKAASTEFKPPGCFMLGBOIYVMOAGTTPMNIFFPISLYLMEVTNQ 120

QY 311 SFRITILPQOYLRAVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYVVDRAKRIGF 370  
 DB 121 SFRITILPQOYLRAVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYVVDRAKRIGF 180

QY 371 AVSACHVHDEFRTAAVEGFYVTLDMEDCGYNIPOTDESTLMTIAYVMAICALEMPLICL 430  
 DB 181 AVSACHVHDEFRTAAVEGFYVTLDMEDCGYNIPOTDESTLMTIAYVMAICALEMPLICL 240

QY 431 MYCOMRCLRCRLOHDDFADDSLTK 456  
 DB 241 MYCOMRCLRCRLOHDDFADDSLTK 266

RESULT 6  
 Q9H2V8 PRELIMINARY; PRT; 439 AA.

AC Q9H2V8  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CDAL3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PHOCHROMOCYTOMA;  
 RA Li Y., Huang Q., Peng, Y., Song H., Yu Y., Xu S., Ren S., Chen Z.,  
 RA Han Z.;  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF212252; AAG41783.1; -.  
 DR HSSP; P00797; ZREN.  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp.3.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 SQ SEQUENCE 439 AA; 48275 MW; 02EC0E0E50F11602 CRC64;

Query Match 48.0%; Score 1160.5; DB 4; Length 439;  
 Best Local Similarity 50.1%; Pred. No. 1.3e-90;  
 Matches 220; Conservative 78; Mismatches 134; Indels 7; Gaps 4;

QY 18 MYDNLGRSGGQYVEMTVGSPPTNLIVDTGSSNFAVGAPHPHRLHRYOQLSSTYR 77  
 DB 1 MYDNLGRSGGQYVEMTVGSPPTNLIVDTGSSNFAVGAPHPHRLHRYOQLSSTYR 60

QY 78 DLKRGVYVPTQKWEGLCTDLVSTPHGPNVYVRANIAITFESDKRFINGSMWEGTILG 137  
 DB 61 SKGFDVYVKTQGSWTFGEVEDLVITIPKGNSTFLVNIAITFESDNFLGIRKMGITLGL 120

QY 138 AYAEIARPDSDLPEFDSLAKQTHVPMLEIQLGAGFPLNOSSEVLASVGSIMIGIDH 197  
 DB 121 AYATLAPSSSLEFPFSLVTOANIRPVEFMQMGAGLPPVAGS---GTNGSLVLGIEP 177

QY 198 SLATGSLMYTPPIRREMYEVIIVVEINGDOLKMDCKEYNKDSIVSGTTLRLPKRV 257  
 DB 178 SLATGSLMYTPPIRREMYEVIIVVEINGDOLKMDCKEYNKDSIVSGTTLRLPKRV 237

QY 258 EAAVSKTAASTEFKPPDGFMLGBOIYVMOAGTTPMNIFFPISLYLMEVTNOSFTTL 317  
 DB 238 DAVEAARASLAPFESDGFWSQLCNSEFPMSYFKISTYLRDENSSSFRITIL 297

QY 318 POOYLRAVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYVVDRAKRIGFAVACHV 377  
 DB 298 POLYIQPMGAGLNY-ECYRFGISPTNALVIGATVMEGFYVJFDRQKRVGFAASPCAE 356

QY 378 HDEFRTAAVEGFYVTLDMEDCGYNIPOTDESTLMTIAYVMAIC-ALFMLPCLMCOMR 436  
 DB 357 IAGAASEISGFSTEDVASNCVPAQSLSEPIIMIVSYALMSVGAILLVILLPLFR 416

QY 437 CLRCLROHDDFADDSL 455  
 DB 417 CLR--RPRDPEVNDSESL 433

RESULT 7  
 Q9JL18 PRELIMINARY; PRT; 514 AA.

AC Q9JL18  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ASPARTYL PROTEASE 1.  
 GN BACE2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choi D.K., Sugano S., Sakaki Y.;  
 RA "Molecular characterization of the mouse Aspl gene, a homolog of the  
 RT human Aspl (Down Syndrome Region aspartyl protease).";  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF216310; AAF36599.1; -.  
 DR HSSP; P32329; 1YPS.  
 DR MEROPS; A01.041; -.  
 DR MGD; MGI:1860440; Bace2.  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp.3.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_2.  
 KW Protease.  
 SQ SEQUENCE 514 AA; 55799 MW; A70725F2C1D5E5B47 CRC64;

Query Match 47.4%; Score 1146.5; DB 11; Length 514;  
 Best Local Similarity 50.7%; Pred. No. 2.6e-89;  
 Matches 216; Conservative 75; Mismatches 130; Indels 5; Gaps 3;

QY 8 EPRGRGSEFVENVDNIRKSGGQYVEMTVGSPPTNLIVDTGSSNFAVGAPHPHRLH 66  
 DB 65 EPRVATANFLAMVDNLOGSGRYLEMLIGPPKQVQLVDTGSSNFAVGAPHSYTD 124

QY 67 YQROQLSSTYRDLKRGVYVPTQKWEGLCTDLVSTPHGPNVYVRANIAITFESDKRFI 126  
 DB 125 YFDSSESSYHSKGFYVYKTYOGSNTGFGEDLVITIPKGFSSPLVNIAITFESDNFL 184

QY 127 NGSNMEGILGLAYAEIARPDSDLPEFDSLAKQTHVPMLEIQLGAGFPLNOSSEVLASV 186  
 DB 185 PGIKMGILGLAYAEIARPDSDLPEFDSLAKQTHVPMLEIQLGAGFPLNOSSEVLASV 246

QY 187 GGSMTIGGIDHSLSYTGSLMYTPPIRREMYEVIIVVEINGDOLKMDCKEYNKDSIVSG 246  
 DB 242 GGSILVIGIEPSLYKGDIVYTPPIKEMYVQIILKLEIGGQNLNDCREYNADKAIYVSG 301

QY 247 TTNLRPLPKRVFAAASKIKAASTEFKPPDGFMLGBOIYVMOAGTTPMNIFFPISLYLME 306  
 DB 302 TTNLRPLPKRVFAAASKIKAASTEFKPPDGFMLGBOIYVMOAGTTPMNIFFPISLYLME 361





```

ID Q9P0D2 PRELIMINARY: PRT: 213 AA.
AC Q9P0D2:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=CORD BLOOD.
RC Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA "Human partial CDS cloned from cd34+ stem cells."
RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
RL EMBL: AF161367; AAF28927.1; -.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp; 1.
FT NON_TER
SQ SEQUENCE 213 AA: 24338 MW: EC9D3FA31CFA835C CRC64;

Query Match 29.5%; Score 712.5; DB 4; Length 213;
Best Local Similarity 83.5%; Pred. No. 7,8e-53;
Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

QY 193 GGDHSLYGLSMTYPTIRREMYEVIIVVEINIGDLMKDCKEYNDKSIIVDSGTTNRL 252
DB 1 GGDHSLYGLSMTYPTIRREMYEVIIVVEINIGDLMKDCKEYNDKSIIVDSGTTNRL 60

QY 253 PKKVEFAAANKSKAASSTKFPDGFGLGOLVCMQAGTTPMNFPIVLSYLMGEVTNOSF 312
DB 61 PKKVEFAAANKSKAASSTKFPDGFGLGOLVCMQAGTTPMNFPIVLSYLMGEVTNOSF 120

QY 313 RTIILPQOYLRPEDVATSQDCYKFAISQSGSTGYMGAVINEG 356
DB 121 RTIILPQOYLRP-----WKMMRPKPTVTVCHLTIVLHG 153

RESULT 11
Q9R1P7 PRELIMINARY: PRT: 255 AA.
AC Q9R1P7:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Accarini M., Fumagalli P., Taramelli R., Ottolenghi S.;
RT "Cloning of a gene from chromosome 21 Down region encoding a potential
transmembrane protease."
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
DE EMBL: AF051150; AAD45964.1; -.
DR MEROPS: A01.041; -.
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp; 2.
DR PRINTS: PR00792; PEPSTN.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
FT NON_TER
SQ SEQUENCE 255 AA: 28685 MW: 53DB317815996D63 CRC64;

Query Match 24.7%; Score 596.5; DB 11; Length 255;
Best Local Similarity 47.8%; Pred. No. 8e-43;

```

```

Matches 109; Conservative 44; Mismatches 74; Indels 1; Gaps 1;

QY 205 WYTIPIREMYEVIIVVEINIGDLMKDCKEYNDKSIIVDSGTTNRLPKKVEFAAANKSI 264
DB 1 WYTIPIREMYEVIIVVEINIGDLMKDCKEYNDKSIIVDSGTTNRLPKKVEFAAANKSI 60

QY 265 KAASSTKFPDGFGLGOLVCMQAGTTPMNFPIVLSYLMGEVTNOSFRITLIPQOYLRP 324
DB 61 ARTSLIPFSGFTGQNLACWTSETPMAFPKISLYLRNENASRSFRITLIPQOYLRP 120

QY 325 VEDVATSQDCYKFAISQSGSTGYMGAVINEGFTYVDRARKRGFAVSACHYDERIA 384
DB 121 MMGAGFNY-ECYRFGISSSTNALVIGATVMEGFYVDRARRGFAVSPAELETGTVS 179

QY 385 AVEGFEVTLMDGCGYNIPQDDESTLMTAYVMAICATLPLPLCLMW 432
DB 180 EISGFESTEDIASNCVPAALNEPLMTLVSTALMSVCAILLVILL 227

RESULT 12
Q9VKP6 PRELIMINARY: PRT: 391 AA.
AC Q9VKP6:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CG17134 PROTEIN.
CN CG17134.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Plankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferrara C., Fertler S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Iibegyan C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jinnai B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

"The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 DR EMBL: AE003630; AAF53016.1; -.  
 DR HSSP: P00794; 4CMS.  
 DR MEROPS: A01.0PW; -.  
 DR F1ybase: FBgn0032304; CG17134.  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp.1.  
 DR PRINTS: PR00792; PERSIN.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 DR Aspartyl protease; Hydrolase.  
 KW SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;

Query Match 13.7%; Score 332.5; DB 5; Length 391;  
 Best Local Similarity 29.1%; Pred. No. 5e-20;  
 Matches 105; Conservative 61; Mismatches 146; Indels 49; Gaps 13;

QY 20 DNLKSGGQYVEMTVGSPQTLNLTVDGSSNFAYGAPHPF-----LHRYQRLS 73  
 DB 66 EHLHSMNNEYGVATIGTPEQRFNILEDTGSANLWBSASCPASNTACQRRNKYDSSAS 125  
 QY 74 STYRLRKGVVVPYTGKWEGLGTLVSIPHPNVTVRANIAITTESDKFINGSNMG 133  
 DB 126 STYVANGEEFAIEYGTSLSGFLSNDIVTIA-GISIQNGTEALSEFGTTFVD-APRAG 183  
 QY 134 ILGLVAELARDDSLPEPFDLSVKQTHVPN-LFSLQLCGAGFPLNQSEVLASVGSMIT 192  
 DB 184 ILGLVFSALIA--VDGVTPEDNMISQGLDEPVIISFYLRKRG-----TAVRGGELIL 233  
 QY 193 GGDHSILGSLMYPIRREWEYEVIIIVREINQGLDKMDCKEYNYDKSIYDSGTNRL 252  
 DB 234 GGIDSLVRSGLTYVPVPAWPKVMTIKTNGTLLCNGC-----QALDTGSLIAV 287  
 QY 233 PKVVEAAVSKSAASSTEKEPFDGFWLGEOLV-CWQAGTTPWNIFFVLSLYLMEVNTQS 311  
 DB 288 PLAAARKINROLGATDND-----GEAFVRCGRVSS-----LPKYNLNIQGV---- 329  
 QY 312 FRITLPOOYLRPVEDVATNSDDCYK-PAISQSSGTVMGAVIMGEFVVEDRARKRGIF 370  
 DB 330 --FTLAPRDYI--KVTONGQYCKSAFTYMEGLSEFWILGDFYIGKFTYVDKNERIGF 385  
 QY 371 A 371  
 DB 386 A 386

RESULT 13  
 Q9GYX7 PRELIMINARY; PRT; 354 AA.  
 AC Q9GYX7;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HEME-BINDING ASPARTIC PROTEINASE (FRAGMENT).  
 OS Boophilus microplus (Cattle tick).  
 CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
 CC Parasitiformes; Ixodidae; Ixodidae; Boophilus.  
 OC NCBI\_Taxid=6941;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PORTO ALEGRE; TISSUE=OVARY;  
 RA Sorgiano M.H.F., Loguillo C., Zingali R.B., Paiva-Silva G.O.,  
 RA Juliano L., Oliveira P.L.;  
 RT "A heme-binding aspartic proteinase from the eggs of the hard tick  
 RT Boophilus microplus".  
 RL J. Biol. Chem. 0:0-0(2000).  
 DR EMBL: AF286865; AAG00993.1; -.  
 DR HSSP: P00797; 2REN.  
 DR MEROPS: A01.054; -.

DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp.1.  
 DR PRINTS: PR00792; PEPsin.  
 FT NON\_TER 1  
 SQ SEQUENCE 354 AA; 38512 MW; B50D7C638CF27091 CRC64;

Query Match 13.7%; Score 332; DB 5; Length 354;  
 Best Local Similarity 26.2%; Pred. No. 4.8e-20;  
 Matches 101; Conservative 74; Mismatches 135; Indels 76; Gaps 15;

QY 19 VDNLRGK-----SGQGYVEMTVGSPQTLNLTVDGSSNFAYGAPHPFL- 64  
 DB 13 VTEIRGRLSPPIPLITNYYNMQFYGITIGTPQSKFLMDGSSNFVW-----PSIN 67  
 QY 65 -----HRYQRLSSTYRLRKGVVVPYTGKWEGLGTLVSIPHPNVTVRANI 115  
 DB 68 CDQSMACRHDHAKYDSKSTFTKSGRYRIRYSGVYRGISIDNVCV--CPATVQYKF 125  
 QY 116 AAITESDKFINGSNWESITLGLAELARPDLSLEPFDLSVKQTHVPN-LFSLQLCGAG 174  
 DB 126 AEMHSDGKLFRRNAKYDIFGLAPPSISQ--NQLPLFDAMVKGQVROAVFSLYL--SK 181  
 QY 175 FPLNQSEVLASVGSMITGIDHSILYTGSLMYTPIRREWEYEVIIIVREINQGLDKM-DC 233  
 DB 182 QPSDGN-----GGEIYFGGINAQRITGATRYVVSQAAMQVMDINVOGTTLCVGGC 235  
 QY 234 KEYNDKSIYDSGTNTNLRPKVFEAAVKSITKASSTEKEPFDGFWLGEOLV-CWQAGTTPW 293  
 DB 236 -----PIYVDSGTSFSLSGP-----SADVETLNRVIGATKTAAGV-----FEVNCATI 277  
 QY 294 NIPVVISLYLME---VTNOSFRITILPOOYLRPVEDVATNSDDCYKPAISQSSGTI--- 347  
 DB 278 SSLPPIITFNLGKSPFIQGEAYTIRI-----PLTGGEGCFTRISEDSASGTNLW 327  
 QY 348 VMGAVIMEGEFVVEDRARKRGIFAVS 373  
 DB 328 ILGAVFTQTYTYVDFRAQNRRGFATA 353

RESULT 14  
 Q9BGUS PRELIMINARY; PRT; 386 AA.  
 AC Q9BGUS;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE CATHEPSIN D (FRAGMENT).  
 GN CAT-D.  
 OS Bos taurus (Bovine).  
 CC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_Taxid=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Higuchi M., Miyashita N., Nagamine Y., Awata T.;  
 RT "Complementary DNA sequence of bovine cathepsin D.";  
 RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC EMBL: AB05312; BAB21620.1; -.  
 DR HSSP: P07339; 1LYB.  
 DR MEROPS: A01.009; -.  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp.1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Aspartyl protease; Hydrolase.  
 FT NON\_TER 1  
 SQ SEQUENCE 386 AA; 41965 MW; 33BF00B80585490D CRC64;

Wed Oct 30 14:11:50 2002

us-09-724-571-43.rsp

Page 8

Query Match	12.9%;	Score 312.5;	DB 6;	Length 386;
Best Local Similarity	28.0%;	Pred. No. 2.5e-18;		
Matches 113;	Conservative	69;	Mismatches 131;	Indels 91;
				Gaps 20;

```

OY      8  EP-GRSRYFVMDNLNRKSGGOCYVVMYVGPSPOLLNLVDFGSSNFMVGAAPHPFL-- 64
Db      35  EPARYQGPPIPELKNKMYDAO--YIGETIGTPPOCTYVFDFGSSANLWNPISLHCKLLDI 91
OY      65  ---HARYQROLSTSYRDLRKGAV--VPYQKGWEGELGTDLSIPHGPN-----VTVR 112
Db      92  ACMTHRKRYNSDKSTY--VANGTFEDLHYGSGSLGTLSDPTVSYVPCNPSSSPGVATVO 149
OY      113 ANI--AAITSEDKFINGSWNEGILGAYELIARPPDSLEPFVDSLYKQTHV-PNFSIQ 169
Db      150  RQTFGEAIKQGVAFI-AAKFGGILGMAYPRIIS--VANNVLPVEDNLMQOOLYDKNWFSS-- 204
OY      170  LCGAGFPLNOSBEVLASVSGSMITGGIDHSILYTSGLTWPTPRRMEYEVIIIVREIINGODL 229
Db      205  ----FLNLR-DPRAQPGGELMGLGDTGSKYKRGSLHNNHTVRAQYIWOIIMDDOLDV-GSSL 257
OY      230  KMDKEYNYDKSIYDSGTTNLRKRYEAAVAKSIKAASSTERKPPDGFWMJGEOLVCMQAG 289
Db      258  TV-CK--GGCAIAYDTGSLVGVVEEVRRLQAL-----G 290
OY      290  TTPMNIFFVLSLYMGEVNTOSFRITLLPQOYLR-PVEDVATSDODCYKRAISQSTGT-- 347
Db      291  AVP-----LIQGEYMIPECEVYSSLPQVYKVLGKRDYKXSPED--YALKVSQAGTTVC 340
OY      348  -----VMGAVIMEGFRVYVEDRRKRIEFGVAS 374
Db      341  LSGEMGMDIPPPGGPLMTLGSVFLIGRTYIVFDNRQONFRGLAEAR 384

```

```

RESULT 15
09DEC2
ID 09DEC2 PRELIMINARY; PRT; 384 AA.
AC 09DEC2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PEPsinogen A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia, Pipidoae, Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasunasu S.;
RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog
RT Rana catesbeiana."
RL Submitted (JUL-2000) to the EMBL/Genbank/DDBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
DR EMBL; AB045380; BAB20798.1; -.
DR HSSP; P00790; 1PSN.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Aspartyl protease; hydrolase.
SQ SEQUENCE 384 AA; 4188 MW; 7290D2E7933D9072 CRC64;

```

Query Match	12.7%	Score 308	DB 13	Length 384
Best Local Similarity	27.5%	Pred. No. 66-18		
Matches	99	Mismatches 133	Indels 68	Gaps 16
30	YVVEMTVGSPPQTLNLIVDTGSSNFVNGAPHFLL-----HRYVQRLSSTYRLRLKRG	82		
11	:::::     ::			

```

Db      72  YGCTISICTPPOEFFTVFIDTSSANLW---PSYVCSSQACSNHNRFPQOSSSFQATNTP 128
QY      83  YVVPYTGKRWGELGTDLVSIPIGPNTVTRANIAITBESK-PIFINGSNMEGILGAYAE 144
Db      129  VSIQGTGSMGSGFGLYDPILOV--GNIOISNOMGSLSESGSFLLYSPDGIILGLAFPS 185
QY      142  IARPDDSLPEPFDSLVKQTHVP-NLEFSQLCGACFPINOSEVLASVGSAMIIGIDHSLY 200
Db      166  IIA--SSQATPVPFDMNWSGGLIPQMLEFVYLLSDG-----QTSYVLEFGVDNSY 233
QY      201  TGSMTYPIRREMYEVIIVAVEINGDL--KMCKEYNDKSTVDSGTTNLRLPKKVE 258
Db      234  SGSLMWVPLAETFYQITLDSVSTINGOVLACQSC-----QALVDTGTSIMTGPSPIT 266
QY      259  AAVSKIKAASTREKPPDFEWMGEOLY-CWAQGTTPWNIFFYISLYLGEVYNOSFRITIL 317
Db      287  ANIQYVIGASQDSN-----GQVYINCNNTSNMPTIIF-----TIN 321
QY      318  PQOY-LREVEDANTQODCKR-FNISQSTGT-----WGAIVMEGFYVEDRARRKIGFA 371
Db      322  GVQYPLSPSAIVKRONQCCSSSFQAMNLPJNSGLMTILGDFINQYTFVDFBRANNYVAA 384

```

Search completed: August 7, 2002, 09:17:12  
Job time: 216 sec

381

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 17.62 Seconds  
(without alignments)  
1002.050 Million cell updates/sec

Title: US-09-724-571-43

Perfect score: 2419  
Sequence: 1 ETDEPEPGRGSGFVEMWD.....CLRLRQHHDFADDISILK 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2419	100.0	501	1	BACE_HUMAN
2	2392	98.9	501	1	BACE_RAT
3	2390	98.8	501	1	BACE_MOUSE
4	1171.5	48.4	518	1	BACE2_HUMAN
5	327	13.5	324	1	PEP1_GADMO
6	314.5	13.0	390	1	CATD_BOVIN
7	306	12.6	387	1	PEP1_RABIT
8	302.5	12.5	388	1	PEP4_MACRU
9	302	12.5	367	1	PEP4_CHICK
10	299.5	12.4	366	1	CATE_HUMAN
11	298.5	12.3	383	1	PEPE_CHICK
12	296	12.2	387	1	PEP2_RABIT
13	296	12.2	391	1	CATE_CAVPO
14	295.5	12.2	412	1	CATD_HUMAN
15	295	12.2	387	1	PEP4_RABIT
16	289.5	12.0	388	1	PEP2_MACRU
17	287	11.9	407	1	CATD_RAT
18	286	11.8	387	1	PEP3_RABIT
19	285.5	11.8	398	1	CATE_RAT
20	285	11.7	398	1	CATD_CHICK
21	283.5	11.7	388	1	PEP4_MACRU
22	283.5	11.6	381	1	CHYM_SHEEP
23	281.5	11.6	388	1	CHYM_SHEEP
24	281.5	11.6	410	1	PEP1_MACRU
25	279.5	11.6	386	1	CATD_MOUSE
26	278.5	11.5	387	1	PEP4_PIG
27	278	11.5	396	1	PEP4_CALA
28	277.5	11.5	397	1	CATE_MOUSE
29	277.5	11.5	397	1	CATE_MOUSE
30	276.5	11.4	419	1	CARV_CAVUL
31	273.5	11.3	381	1	CHYM_BOVIN
32	273	11.3	388	1	PEP4_RABIT
33	270.5	11.2	377	1	PEP4_MACRU

## ALIGNMENTS

RESULT ID	1	BACE_HUMAN	STANDARD	PRT	501 AA
AC	P56817	Q9UT5			
DT	30-MAY-2000	(Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)				
DE	(Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2) (Memapsin-2).				
DE	BACE OR BACE1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxId=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM A).				
RC	TISSUE=Brain;				
RX	MEDLINE=20002972; PubMed=10531052;				
RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A., Denis P., Teplow D.B., Ross S., Amaratne P., Loefloft R., Luo Y., Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A., Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F., Treanor J., Rogers G., Citron M.;				
RT	"Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";				
RL	Science 286:735-741(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND CHARACTERIZATION.				
RC	TISSUE=Brain;				
RX	MEDLINE=20057171; PubMed=10591214;				
RA	Silva S., Anderson J.P., Barbour R., Basl G.S., Caccavello R., Davis D., Dean M., Dovey H.F., Frigon N., Hong J., Jacobson-Croak K., Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H., Taksuno G., Tung J., Schenk D., Seubert P., Sismonsaari S.M., Wang S., Walker D., Zhao J., McConlogue L., Varghese J.;				
RT	"Purification and cloning of amyloid precursor protein beta-secretase from human brain.";				
RL	Nature 402:537-540(1999).				
RN	[3]				
RP	SEQUENCE FROM N.A. (ISOFORM A).				
RX	MEDLINE=20057170; PubMed=10591213;				
RA	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M., Brasher J.R., Strahan N.C., Mathews W.R., Buhl A.E., Carter D.B., Tomasselli A.G., Parodi L.A., Helinkson R.L., Gurney M.E.;				
RT	"Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity.";				
RL	Nature 402:533-537(1999).				
RN	[4]				
RP	SEQUENCE FROM N.A. (ISOFORM A).				
RX	MEDLINE=20030166; PubMed=10561122;				
RA	Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D., Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M., Smith T.S., Simons D.L., Walsh F.S., Dingwall C., Christie G.;				

RT "Identification of a novel aspartic proteinase (Asp 2) as beta-secretase.";  
 RT Mol. Cell. Neurosci. 14:419-427(1999).  
 RL [5]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Brain, and Pancreas;  
 RA Michel B., De Pietri Tonelli D., Zacchetti D., Keller P.;  
 RT "New beta-site APP cleaving enzyme isoform (BACE-1B) obtained from human brain and pancreas.";  
 RL Submitted (Jan-2001) to the EMBL/Genbank/DBJ databases.  
 RP [6]  
 RP SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.  
 RA MEDLINE=20144060; PubMed=10677483;  
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;  
 RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP, LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A/BACE-1A (shown here) and B/BACE-1B; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AF190725; AAF04142.1; -;  
 DR EMBL; AF201468; AAF18982.1; -;  
 DR EMBL; AF200343; AAF17079.1; -;  
 DR EMBL; AF204943; AAF26367.1; -;  
 DR EMBL; AF338816; AAK38374.1; -;  
 DR EMBL; AF200193; AAF13715.1; -;  
 DR HSSP; P32329; 1YPS.  
 DR MIM; 604252; -;  
 DR InterPro; IPR001969; Asp\_protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp; 3.  
 DR PRINTS; PR00792; PEPsin.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolyase; Aspartyl protease; Glycoprotein; zymogen; Transmembrane;  
 KW Signal; Alternative splicing.  
 FT SIGNAL 1 21  
 FT PROPEP 22 45  
 FT CHAIN 46 501  
 FT DOMAIN 22 457  
 FT TRANSMEM 458 478  
 FT DOMAIN 479 501  
 FT ACT\_SITE 93 93  
 FT ACT\_SITE 289 289  
 FT CARBOHYD 153 153  
 FT CARBOHYD 172 172  
 FT CARBOHYD 223 223  
 FT CARBOHYD 354 354  
 FT CARBOHYD 190 214  
 FT VARSPLIT 501 AA; 55763 MM; 377CE4C824ACEF05 CRC64;  
 SQ

Query Match 100.0%; Score 2419; DB 1; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-193;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEPRGRGSFVEMVDNLRGKSGGYVEMTVGSPQTLIIIVDTGSSNFAVGAAP 60  
 Db 46 ETDEPEPRGRGSFVEMVDNLRGKSGGYVEMTVGSPQTLIIIVDTGSSNFAVGAAP 105  
 QY 61 HPFLHRYQROLSSTYRDLRKGVVYPYQGMKEGELDTLVSIHPGNVYRANIAATE 120  
 Db 106 HPFLHRYQROLSSTYRDLRKGVVYPYQGMKEGELDTLVSIHPGNVYRANIAATE 165  
 QY 121 SDKEFINSNMEGILGLAVETARPDDSLPEFPISLVKQTHVPMPLFSLQCGAFPLNOS 180  
 Db 166 SDKEFINSNMEGILGLAVETARPDDSLPEFPISLVKQTHVPMPLFSLQCGAFPLNOS 225  
 QY 181 EYLVSGVSMIIIGIDHSLYGLSLWYTPIRREWYEVITVVEINIGODLKMCKEYNDK 240  
 Db 226 EYLVSGVSMIIIGIDHSLYGLSLWYTPIRREWYEVITVVEINIGODLKMCKEYNDK 285  
 QY 241 STVDGTTNLRPKKVFEEAAVKSIAKASTKFPDGFVLSQVLCWAGTTPNNIFPVIS 300  
 Db 286 STVDGTTNLRPKKVFEEAAVKSIAKASTKFPDGFVLSQVLCWAGTTPNNIFPVIS 345  
 QY 301 LYLMEVTNOSRRTIILPQYLRPVEDVATSDODCYKFAISOSTGTVMGAVIMEGFYV 360  
 Db 346 LYLMEVTNOSRRTIILPQYLRPVEDVATSDODCYKFAISOSTGTVMGAVIMEGFYV 405  
 QY 361 FDRARRIGFAYASACHVDEPRTAAVEGPFYLLDMEDGYNIPQDESTLMTIAYVNAI 420  
 Db 406 FDRARRIGFAYASACHVDEPRTAAVEGPFYLLDMEDGYNIPQDESTLMTIAYVNAI 465  
 QY 421 CALFMLPLCLMWQWRCLRCRQOHDFADISLILK 456  
 Db 466 CALFMLPLCLMWQWRCLRCRQOHDFADISLILK 501  
 RESULT 2  
 BACE\_RAT ID BACE\_RAT STANDARD: PRT: 501 AA.  
 AC P56819;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
 DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
 DE protease 2) (Asp 2) (Asp2) (Membrane-associated aspartic protease 2)  
 DE (Memapsin-2).  
 GN BACE.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20002972; PubMed=10531052;  
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
 RA Bieri A.L., Curran E., Burgess T., Louis J.C., Collins F.,  
 RA Treanor J., Rogers G., Citron M.;  
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";  
 RL Science 286:735-741(1999).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP, LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF190727; AAF04144.1; -.  
 DR HSSP: P32329; 1YPS.  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp; 3.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KM Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
 KW Signal.  
 FT SIGNAL. 1 21 POTENTIAL.  
 FT PROPEP 22 45 POTENTIAL.  
 FT CHAIN 46 501 BETA-SECRETASE.  
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 458 478 POTENTIAL.  
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 289 289 BY SIMILARITY.  
 FT CARBOHD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 501 AA; 55806 MW; 24B445BC8BE87DE3 CRC64;

Query Match 98.9%; Score 2392; DB 1; Length 501;  
 Best Local Similarity 98.2%; Pred. No. 9.2e-191;  
 Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGRSGFVENVNDLRKSGGGYVEMTGSPPQTLNLTIVDTGSSNFAVGAAP 60  
 DB 46 ETDEPEEPGRGRSGFVENVNDLRKSGGGYVEMTGSPPQTLNLTIVDTGSSNFAVGAAP 105  
 QY 61 HPELRRYQROLSTYRDLRKGVYPRYQKGEGELGTDLVSIHPGPNVYRANATAITE 120  
 DB 106 HPELRRYQROLSTYRDLRKGVYPRYQKGEGELGTDLVSIHPGPNVYRANATAITE 165  
 QY 121 SDKFFINSNMEGILGLVAETARPDDSLPEPDSLVKOTHPNLFSLQDLCAGGFPNLOS 180  
 DB 166 SDKFFINSNMEGILGLVAETARPDDSLPEPDSLVKOTHPNLFSLQDLCAGGFPNLOS 225  
 QY 181 EVLASVSGSMITGGIDSLYTGSLWYTPIRREMYEVIIVREINGODLKMCKEYNYDK 240  
 DB 226 EVLASVSGSMITGGIDSLYTGSLWYTPIRREMYEVIIVREINGODLKMCKEYNYDK 285  
 QY 241 SIYDSGTINLRPKKVFENAVKSTKAASSTKFPDGFVLGEOLVCWQAGTTPWNIFFVIS 300  
 DB 286 SIYDSGTINLRPKKVFENAVKSTKAASSTKFPDGFVLGEOLVCWQAGTTPWNIFFVIS 345  
 QY 301 LYLMGEVTSNOSFRITILPOQYLRPEVATSDODCYKFAISOSSTGTWGVIMGEFVV 360  
 DB 346 LYLMGEVTSNOSFRITILPOQYLRPEVATSDODCYKFAISOSSTGTWGVIMGEFVV 405  
 QY 361 FDRARRKRGFAVSAHVDEFTAAVEGPFYTLDMEDGYNIPQDESTLMTIAVYMAAI 420  
 DB 406 FDRARRKRGFAVSAHVDEFTAAVEGPFYTLDMEDGYNIPQDESTLMTIAVYMAAI 465  
 QY 421 CALFPLPLCLAVWCQRCLRCRQOHDDPADDISLTK 456  
 DB 466 CALFPLPLCLAVWCQRCLRCRQOHDDPADDISLTK 501

RESULT 3  
 ID BACE\_MOUSE STANDARD; PRT; 501 AA.  
 AC P56818;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
 DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
 DE protease 2) (Asp 2) (Asp2) (Membrane-associated aspartic protease 2)  
 DE (Memapsin 2).  
 GN BACE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20002972; PubMed=10531052;  
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeflof R., Luo Y.,  
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
 RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
 RA Treanor J., Rogers G., Citron M.;  
 RT "beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
 RT the transmembrane aspartic protease BACE.";  
 RL Science 286:735-741(1999).  
 RN [2]  
 RP REVISIONS TO 6 AND 81-87.  
 RA Bennett B.D., Vassar R., Citron M.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RP [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057170; PubMed=10591213;  
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,  
 RA Brashler J.R., Strachan N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
 RA Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gunney M.E.;  
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease  
 RT beta-secretase activity.";  
 RL Nature 402:533-537(1999).  
 CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
 CC FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: BRAIN.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF190726; AAF04143.2; -.  
 DR EMBL: AF200346; AAF17082.1; -.  
 DR HSSP: P56272; 1AM5.  
 DR MGD: MGI:1346542; Bace.  
 DR InterPro: IPR001969; Asp.protease.  
 DR InterPro: IPR001461; pepsin.  
 DR Pfam: PF00026; asp; 1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
 KW Signal.  
 FT SIGNAL. 1 21 POTENTIAL.  
 FT PROPEP 22 45 POTENTIAL.  
 FT CHAIN 46 501 BETA-SECRETASE.  
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 458 478 POTENTIAL.  
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT ACT\_SITE 289 289 BY SIMILARITY.



```

FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 501 AA; 55747 MM; C085A013145E474E CRC64;

Query Match 98.8%; Score 2390; DB 1; Length 501;
Best Local Similarity 98.2%; Pred. No. 1.3e-190;
Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 EDDPEEPGRGSRGSEVEMVDNLKRGSGGGYEMTVGSPPTLNLIVDTGSSNFAVGAAP 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 46 ETDSESEBGRGSRGSEVEMVDNLKRGSGGGYEMTVGSPPTLNLIVDTGSSNFAVGAAP 105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 61 HPELHRYQROLSTYRDLRKGVVYPTQGWGELGTDLVSIIPHGPNTVYRANIAAITE 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 106 HPELHRYQROLSTYRDLRKGVVYPTQGWGELGTDLVSIIPHGPNTVYRANIAAITE 165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 121 SDFKFTNGSNMGITIGLAAELIARPDSDLPEPDSLVKOTHPNLPSIOLGAGFPLNOS 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 166 SDFKFTNGSNMGITIGLAAELIARPDSDLPEPDSLVKOTHPNLPSIOLGAGFPLNOS 225
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 181 EYLAIVGSGMIIIGIDHSIYTGSLWYTPIRREMYEVIIVREINGODLKMCKEYNDK 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 226 EYLAIVGSGMIIIGIDHSIYTGSLWYTPIRREMYEVIIVREINGODLKMCKEYNDK 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 241 STVDSGTTNLRIPKRYFEAAVKSIAASTEKPPDGFMLGEOVLCMOAGTTWNTFFPVIS 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 286 STVDSGTTNLRIPKRYFEAAVKSIAASTEKPPDGFMLGEOVLCMOAGTTWNTFFPVIS 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 301 LYLMGEVTVQSRITITLPOQYLRPEVDVATSDDDCYKFAISOSSTGTVGAVIMEGFYV 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 346 LYLMGEVTVQSRITITLPOQYLRPEVDVATSDDDCYKFAISOSSTGTVGAVIMEGFYV 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYINIPQDESLTMTIAYMAAI 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 406 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYINIPQDESLTMTIAYMAAI 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 421 CALFMLPCLCMLVCOMRCRLCROOHDFADDSLILK 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 466 CALFMLPCLCMLVCOMRCRLCROOHDFADDSLILK 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
BAE2_HUMAN STANDARD: PRT; 518 AA.
ID BAE2_HUMAN STANDARD: PRT; 518 AA.
AC Q9Y5Z0; Q9YJTG; 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving
DE enzyme 2) (Aspartyl protease 1) (Asp 1) (Asp1) (Membrane-associated
DE aspartyl protease 1) (Memapsin-1).
GN BACE2 OR ASP21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=2005170; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;
RT "Membrane-anchored aspartyl protease with Alzheimer's disease
RT beta-secretase activity.";
RL Nature 402:533-537(1999).
RN [2]
RN SEQUENCE FROM N.A.
RX TISSUE=bone marrow;
RA Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,
RA Giese K.;

```

```

RT "Identification of a novel aspartic-like protease differentially
RT expressed in human breast cancer cell lines.";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RA Acciarino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;
RT "Cloning of a gene from chromosome 21 Down region encoding a potential
RT transmembrane aspartyl protease.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RX MEDLINE=20030166; PubMed=10561122;
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,
RA Smith T.S., Simons D.L., Walsh F.S., Dingwall C., Christie G.;
RT "Identification of a novel aspartic proteinase (Asp 2) as
RT beta-secretase.";
RL Mol. Cell. Neurosci. 14:419-427(1999).
RN [6]
RN SEQUENCE FROM N.A.
RX MEDLINE=20144060; PubMed=10677483;
RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;
RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of
RT beta-amyloid precursor protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).
RN [7]
RN SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Onk K., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Leimann R., Patterson D.,
RA Reichwald A., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: AF200342; AAF17078.1; -
DR EMBL: AF117892; AAD45240.1; -
DR EMBL: AF050171; AAD45963.1; -
DR EMBL: AF178532; AAF29494.1; -
DR EMBL: AF204944; AAF26368.1; -
DR EMBL: AF200192; AAF13714.1; -
DR EMBL: AL163284; CAB90458.1; -
DR EMBL: AL163285; CAB90554.1; -
DR HSSP: P00797; 2REN.
DR MIM: 605668; -
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001461; Pepsin.
DR Pfam: PF00026; asp; 3.
DR PRINTS: PR00792; PEPsin.

```



```

RT cod (Gadus morhua) ".":
RL Acta Crystallogr. D 54:32-46(1998).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
DR EUKARYOTIC ASPARTYL PROTEASES FAMILY.
PDB: 1AM5: 24-DEC-97.
DR InterPro: IPR001969; Asp-protease.
DR InterPro: IPR001461; pepsin.
DR Pfam: PF00026; asp. 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTEASE: 2.
KW Hydrolase: Aspartyl protease; Digestion; Stomach; Gastric juice;
KW 3D-structure.
FT ACT_SITE 32 32 BY SIMILARITY.
FT ACT_SITE 214 214 BY SIMILARITY.
FT DISULFID 45 50 BY SIMILARITY.
FT DISULFID 206 209 BY SIMILARITY.
FT DISULFID 247 280 BY SIMILARITY.
SQ SEQUENCE 324 AA; 34014 MW; E33A6097B6941DD7 CRC64;

Query Match 13.5%, Score 327; DB 1; Length 324;
Best Local Similarity 27.9%, Pred. No. 9,1e-20;
Matches 104; Conservative 66; Mismatches 137; Indels 66; Gaps 15;

QY 18 MYDNLKRGSGGGYYEMTVGSPPTQLNLVDTGSSNFRVY----AAHPFLHRYQRLS 73
DB 2 VTEQKNKEADLETETGVISIGTPPESEFKYIFDTGSSNMLVSSSHCSAQCNSHNKFKPRQS 61
QY 74 STYRDLRKGVVYVYQGMKEGELGTDLSIPHG--PNVTYANIAATIESDKPFINSNW 131
DB 62 SYETVEKGTVDLTDTGTGMKRGILGDDTVYVSGGSDPNQELG---ESQTECPQPA-AAPF 117
QY 132 EETLLAATAELARPDSLEFPFDSLVKOTHY-PULFSIQLGAGEPINOSEVLASVGS 190
DB 118 DQELGLAPSTIAA--GAVPEFDMNGSGLVEKDLFSYLSGGG--ANGSEVM----- 166
QY 191 IIGGIDHSLYGLSLWYPTPIRREMYEYEVIVVEINCOGLKMD-CKEYNYKSIYDSGTN 249
DB 167 -LGVDSNHYTGSIMHIVTAEKTYQVALDSITVNGQTACCEG-----QALVDIGTSK 219
QY 250 LRLPKVFEAAVKSIAASSTKEFPDFGWLGEQLVCWQAGTTPWNIPVISTLYLMGEVN 309
DB 220 IVAPYSALANIMKLDIGASEN-----QGEHMGN---CASVSLPITP-----TI 260
QY 310 QSEFTTILPQGYLRLRVEVATSDQCYKFAISQSTGT-----VMGAVIMEGLVVF 361
DB 261 NGVKQPLPSPAYIEGDOAFTS-----GLSSGVPNSTSELMIFGVFLNRYTIV 311
QY 362 DRARRKRGFAVSA 374
DB 312 DRTNKKVGFAPAA 324

RESULT 6
CATD_BOVIN STANDARD; PRT; 390 AA.
ID CATD_BOVIN
AC P80209; O9TS27;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Catchpsin D precursor (EC 3.4.23.5).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RX [1]
RN SEQUENCE OF 1-48.
RP TISSUE=Milk;
RX MEDLINE=93202276; Pubmed=8454061;
RA Larsen L.B., Boisen A., Petersen T.E.;
RT "Prochapsin D cannot autoactivate to cathepsin D at 25°C in 1."

```

```

RL  FRBS Lett. 319:54-58(1993).
RN  [2]
RP  SEQUENCE OF 45-390, AND X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RC  TISSUE-Liver;
RX  MEDLINE=93223670; PubMed=8467789;
RA  Metcalf P., Fusek M.;
RT  "Two crystal structures for cathepsin D: the lysosomal targeting
RT  signal and active site.";
RL  EMO J. 12:1293-1302(1993).
CC  -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
CC  PROTEIN BREAKDOWN.
CC  -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
CC  that of pepsin A. Does not cleave the 4-Gln-1-His-5 bond in B
CC  chain of insulin.
CC  -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC  -1- SUBCELLULAR LOCATION: Lysosomal.
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC  EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC  PIR: S32383; S32383.
CC  PIR: S37419; S37419.
CC  HSSP: P07339; 1LVB.
CC  MEROPS: A01.009; -.
CC  InterPro: IPR001969; Asp_protease.
CC  InterPro: IPR001461; Pepsin.
CC  Pfam: PF00026; asp. 1.
CC  PRINTS: PR00141; ASP_PROTEASE. 2
CC  PROSITE: PS00141; ASP_PROTEASE. 2
CC  HydroLase: Aspartyl protease; Glycoprotein; Lysosome; Zymogen.
CC  PROPEP 1 44
CC  CHAIN 45 390
CC  ACT_SITE 77 77
CC  ACT_SITE 273 273
CC  DISULFID 90 97
CC  DISULFID 264 268
CC  DISULFID 307 344
CC  CARBOHD 114 114
CC  CARBOHD 241 241
CC  SEQUENCE 390 AA; 42488 MW; 5B38AA1C33C48D35 CRC64;

Query Match 13.0%; Score 314.5; DB 1; Length 390;
Best Local Similarity 28.0%; Pred. No. 1.3e-18;
Matches 113; Conservative 72; Mismatches 128; Indels 91; Gaps 21;

OY 8 ERG-RGSGVEVMDNRKSGSGGYVEVMVSPPTLNILIVDSSNFAVGAHPHFL-- 64
DB 39 EEPVNRGGPIPELKNYMDAO---YIGELIGITPPQCFYVFDGSAANLWVPSHCKLDI 95
OY 65 ----HRYRQLSTYRDLRKGVY--VPYQKMEGELGTDIVSIPHPN-----VTVR 112
DB 96 ACWTHRKYSNDSKSY--VKNGTTFDIHVGSGSLGYLSQDVTVPNCPSSSSPGCVTVQ 153
OY 113 ANI--AAITESDKFPIINGSWMEGILGLAYAEIARPDSDLEFPDLSLVKQTHV-PNLSLD 169
DB 154 RQTFGEIKRPGVVF-AAKFDDIIGMAYPRIS--VNNVLPVDDNMQQLVAVKNFVS-- 208
OY 170 LCGAGPPLNSEVLASVSGMIIGIDHSLYTGSLSLWTPPIRREMYEVLIVREINGODL 229
DB 209 -----FPLNR-DKAPQGGMLMGIDGSKYKRSGLMFHNVTROAYMG IHMDOLDV-GSSL 261
OY 230 KMDCKEYNVDKSIIVDGGTTLRLPKVFEAAVKSITAASTKEFPDGFVLGEOLV-CWQA 288
DB 262 TV-CK--GCGEALIVDGTSLIVGVEVEERLQKAIGAVPLIO-----GEYNIPEKV 310
OY 289 GTTPWNIFFPISLYLMEVINOSEFRITLIPQQLRPVEDVATSDODCKFAISOSSTGT- 347
DB 311 SS-----LPEVTYKLG-----KDYALSPED-YALKVSOAETTV 344
OY 348 -----VMGAVIMEGFYVVDRAKRIRIGFAVSA 374
DB 345 LSGFMGMDIPPGGFLMILGDVFIKRIYTVFDNRQNVGLAEAA 388

```

```

RESULT 7
PEP1_RABIT
ID PEP1_RABIT STANDARD; PRT; 387 AA.
AC P28712;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pepsin II-1 precursor (EC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -1- DEVELOPMENTAL STAGE: PEPINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC PIR: B38302; B38302.
CC HSSP: P00791; 1PSA.
CC MEROPS: A01.001; -.
CC InterPro: IPR001969; Asp_protease.
CC InterPro: IPR001461; Pepsin.
CC Pfam: PF00026; asp. 1.
CC PRINTS: PR00792; Pepsin.
CC PROSITE: PS00141; ASP_PROTEASE. 2
CC HydroLase: Aspartyl protease; Digestion; Stomach; Gastric juice;
CC zymogen; signal; Phosphorylation; Multigene family.
CC SIGNAL 1 15
CC PROPEP 16 59
CC CHAIN 60 387
CC MOD_RES 129 129
CC ACT_SITE 93 93
CC ACT_SITE 276 276
CC DISULFID 106 111
CC DISULFID 267 271
CC DISULFID 310 343
CC SEQUENCE 387 AA; 42070 MW; A6EC4BF715541A48 CRC64;

Query Match 12.6%; Score 306; DB 1; Length 387;
Best Local Similarity 27.1%; Pred. No. 6.4e-18;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;

OY 30 YVEEMTVGSPPTLNILIVDSSNFAVG-----AAPRFLHRYQRLDLSYRDLRGVYV 85
DB 75 YFGTISIGPPEDETYVIFDGSNLMWVPSSTYCSLACFLHRRNPDDSTFOATSETLSI 134
OY 86 PYQKMEGELGTDIVSIPHPNVTVRANIAITESD--KPIINGSNMEGILGLAYAEI 142
DB 135 TYTGSGSTGLGIDYTKV---GNIEDTINQFGLSKTERPGTFLV--APFGILGLAYPSI 189
OY 143 ARPDSDLEFPDLSLVKQTHV-PNLSLQCGAGPPLNOSSEVLASVSGSMIGIDHSITV 201
DB 190 SASDAT--PVFDNMWMMNGIIVSEDLFSYVLSNG-----EKSQMWFGGIDSSYTV 237
OY 202 GSIWTPPIRREMYEVLIVREINGODLKM--DCKEYNVDKSIIVDGGTTLRLPKVFEA 259
DB 238 GSIWVAVSHSGYWOITMDSITINGETIACADSC-----QAVVDGTSLAAPTSAISK 291

```

QY 260 AVSIRKASSTKPPDFWLGEOIV-CWQAGTTPWNIFFVISTYLKMGVNTQSRITILP 318  
DB 292 IQSYTGASKNL-----LGENIISCSAIDSLPDIYF-----TINN 325  
QY 319 QQTLRPVED-VATSDODC---YKFAISQSSTGT--VWGAIVMEGFVYEDRARRKRGIFAV 372  
DB 326 VQYPLPASAIILKEDDDCLSGFDGMNLDITSYGELWILGDVFIROYFTVFDRAANNQVGLAA 385  
QY 373 SA 374  
DB 386 AA 387  
RESULT 8  
PEP4\_MAFU STANDARD; PRT: 388 AA.  
ID PEP4\_MAFU  
AC P27678;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Pepsin A-4 precursor (EC 3.4.23.1) (Pepsin I/II).  
GN PGA.  
OS Macaca fuscata fuscata (Japanese macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Macaca.  
NCBI\_TaxID=9543;  
[1]  
RN SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.  
RP TISSUE=gastric mucosa;  
RC MEDLINE=92037645; PubMed=1935977;  
RA Kagiyama T., Tanabe K., Kolwai O.;  
RT "Development-dependent expression of isozymogens of monkey  
pepsinogens and structural differences between them."  
RL Eur. J. Biochem. 202:205-215(1991).  
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
ALSO CLEAVED TO SOME EXTENT.  
CC -1- CATALYTIC ACTIVITY: PREPERENTIAL CLEAVAGE PHE-, LEU-.  
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
HORMONES AND RELATED SUBSTANCES.  
CC -1- MISCELLANEOUS: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING  
PEPSIN AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA  
ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC  
CLEAVAGE VIA AN INTERMEDIATE FORM(S).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X59753; CAA42425.1; -  
DR PIR: S16065; S16065.  
DR PIR: S19682; S19682.  
DR HSSP: P00790; IPSN.  
DR MEROPS: A01.001; -;  
DR InterPro: IPR001969; Asp\_protease.  
DR InterPro: IPR001791; Laminin\_G.  
DR Pfam: PF00026; asp.1.  
DR PRINTS: PF00792; PEPsin.  
DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW Zymogen; Multigene family; Signal; Glycoprotein.  
FT SIGNAL 1 15 BY SIMILARITY  
FT PROPEP 16 38 ACTIVATION PEPTIDE.  
FT PROPEP 39 62 ACTIVATION PEPTIDE.  
FT CHAIN 63 388 PEPSIN A-4.

FT ACT\_SITE 94 94 BY SIMILARITY.  
FT ACT\_SITE 277 277 BY SIMILARITY.  
FT DISULFID 107 112 BY SIMILARITY.  
FT DISULFID 268 272 BY SIMILARITY.  
FT DISULFID 311 344 BY SIMILARITY.  
FT CARBOHYD 88 88 N-LINKED (GLCNAC...);  
SQ SEQUENCE 388 AA; 41955 MW; A2923AB1F7FCDEB9 CRC64;  
Query Match 12.5%; Score 302.5; DB 1; Length 388;  
Best Local Similarity 27.6%; Pred. No. 1.3e-17;  
Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;  
QY 3 DEEPPEGRGRGFVEMVDNLKRGSSGGYVEMTVGSPPTGLNLDVTGSSNFAVGAAPHP 62  
DB 64 DEQPLE-----NLDV-----EYFGTIGIGTPAONFTVVEPTGSSNLMV--PSV 105  
QY 63 FL-----HRYVQRLSSTYRDLRKGVVVPYTOGKMEGELTDIVSIPHGNNVVRANI 115  
DB 106 YCVSLACMDHNLFPQDSSYVATSKVTSITGTGSMGILGYDYKV--GGISPTNQT 162  
QY 116 AATESDK-FETNGSNMWSGILGLAVAEIARPDLSLEPFDSLVKQTHV-PNLFSQLCGA 173  
DB 163 FGLSETEPEGFIFYFAFDFDILGLAVPSIS--SSGATPVPDNINQRLVQSDFSVYLSAD 220  
QY 174 GFLNLOSEVLASVSGSMILGIDHSLYTSLWYTPIRREMYEVILVREINQDL--KM 231  
DB 221 ----DOS-----GSVYLFSGIDSSYTGSLNWPVSVGVQWISVDSITNMGKTACAK 270  
QY 232 DKEKNYDKSIVDSGTTNLRPKVFEAAVKSIRKASSTKPPDFWLGEOIV-CWQAGT 290  
DB 271 GC-----QAIYDGTGSLTGTPTSPANIQSDIGASEND-----GEMVSSSAISS 316  
QY 291 TFWNIFFVISTYLKMGVNTQSRITILPQY-LRVEDVATSDQYK-----FAISQSS 344  
DB 317 LDDIYF-----TINGVQYPLPPSAVILYLOSQSGCTSGFQGMVDYPTSG 358  
QY 345 TGTGAVVIMEGFVYEDRARRKIGFA 371  
DB 359 ELWILGDVFIROYFTVFDRAANNQVGLA 385  
RESULT 9  
PEP4\_CHICK STANDARD; PRT: 367 AA.  
ID PEP4\_CHICK  
AC P00793;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Pepsin A precursor (EC 3.4.23.1).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
CC Gallus.  
NCBI\_TaxID=9031;  
[1]  
RN SEQUENCE.  
RP MEDLINE=84004412; PubMed=6617663;  
RA Baudys M., Kostka V.;  
RT "Covalent structure of chicken pepsinogen.";  
RL Eur. J. Biochem. 136:89-99(1983).  
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
ALSO CLEAVED TO SOME EXTENT.  
CC -1- CATALYTIC ACTIVITY: PREPERENTIAL CLEAVAGE PHE-, LEU-.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
DR PIR: A00984; PECH.  
DR HSSP: P00794; 4CMS.  
DR MEROPS: A01.001; -;  
DR InterPro: IPR001969; Asp\_protease.  
DR InterPro: IPR001461; pepsin.  
DR Pfam: PF00026; asp.1.

DR PRINTS: PRO0792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolyase; Aspartyl protease; Digestion; Stomach; Zymogen;  
 KW Glycoprotein; Gastric juice.  
 FT PROPEP 1 367 ACTIVATION PEPTIDE.  
 FT CHAIN 43 367 PEPsin A.  
 FT ACT\_SITE 77 77  
 FT ACT\_SITE 260 260  
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).  
 FT DISULFID 90 95  
 FT DISULFID 251 255  
 FT DISULFID 290 323  
 FT DISULFID 367 367  
 FT DISULFID 404 31 MM; 0C547E7FD8F5B341 CRC64;  
 SO SEQUENCE 367 AA; 404 31 MM; 0C547E7FD8F5B341 CRC64;  
 Query Match 12.5%; Score 302; DB 1; Length 367;  
 Best Local Similarity 24.0%; Pred. No. 1.3e-17;  
 Matches 88; Conservative 69; Mismatches 125; Indels 84; Gaps 13;  
 QY 30 YVEMTVGSPPTLILVDTGSSNFAVGAAPFL-----HRYQRLSTYRDLKRG 82  
 DB 59 YGTFISIGTPQDDFSYIEDTGSSNLW---PSIYKSSACSNHKKRDPKSSYVSTNET 115  
 QY 83 YVVPITQGMKEELGDIIVSIPHPVIVRANIIAITEGDK-FEINGSMGILGLAYAE 141  
 DB 116 VYIAIGTSMGSLIGDYIVAV---SIDVQNOIFGLSETPGSFYCYCFDILGLAFPS 172  
 QY 142 IARPDSLPFFDSLVKQTHV-PNLFSLQCGAGFPLNSEVLAVSGSMITGIDHSILY 200  
 DB 173 IS--SSGAPFVDDNMMSQHLVAODLFVYLSKRG-----ENSGVILFGGIDPNYT 220  
 QY 201 TGSIMWTPFIREYXVYVILVREINGODLK--MDCKEYNYKSIYVSGTNTLRPKVE 258  
 DB 221 TGTIYVWPLSAETIWOITMDRYTVGKKYVACFTC---QAIYVGTSLVMPGAIN 274  
 QY 259 AAVKSIKAASSTE-----KPPDGFMLEQOLVQWAGTTPMNIFFVSLYLMEVTNOS 311  
 DB 275 RIKIDGVSSDDEISCDISKLPD-----VYFHINGHA----- 307  
 QY 312 FRITILPQOYLRVEDVAVTSQDCYFAISQSSTG-----VMGAVIMEGFVVDRAR 365  
 DB 308 -----FTLPASAVYVINEDESCMLGFENMCTPELGEOWILGIVFREYVITDRAN 358  
 QY 366 KRIGFA 371  
 DB 359 NKVGLS 364  
 RESULT 10  
 CATE\_HUMAN STANDARD; PRT; 396 AA.  
 ID CATE\_HUMAN STANDARD; PRT; 396 AA.  
 AC P14091;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Cathepsin E precursor (EC 3.4.23.34).  
 GN CTSE  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89380302; PubMed=2674141;  
 RA Akuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;  
 RT Human gastric cathepsin E. Predicted sequence, localization to  
 RT chromosome 1, and sequence homology with other aspartic  
 RT proteinases.;  
 RL J. Biol. Chem. 264:16748-16753(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92112877; PubMed=1370478;  
 RA Azum T., Liu W.G., Vander Jaan D.J., Bowcock A.M., Taggart R.T.;

RT "Human gastric cathepsin E gene. Multiple transcripts result from  
 RT alternative polyadenylation of the primary transcripts of a single  
 RT gene locus at 1q31-q32.";  
 RL J. Biol. Chem. 267:1609-1614(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Taitnell P.J., Kay J.;  
 RL "Human procathepsin E";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.  
 RX MEDLINE=90241267; PubMed=2334440;  
 RA Athauda S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;  
 RT "Structural evidence for two isozymic forms and the carbohydrate  
 RT attachment site of human gastric cathepsin E";  
 RL Biochem. Biophys. Res. Commun. 168:878-885(1990).  
 CC -1 FUNCTION: DUE OF ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN  
 CC LYMPOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.  
 CC -1 CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader  
 CC specificity.  
 CC -1 SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M84424; AAA52300.1; JOINED.  
 DR EMBL; M84413; AAA52300.1; JOINED.  
 DR EMBL; M84417; AAA52300.1; JOINED.  
 DR EMBL; M84419; AAA52300.1; JOINED.  
 DR EMBL; M84420; AAA52300.1; JOINED.  
 DR EMBL; M84421; AAA52300.1; JOINED.  
 DR EMBL; M84422; AAA52300.1; JOINED.  
 DR EMBL; J05036; AAA52130.1; JOINED.  
 DR EMBL; A750717; CAB82850.1; JOINED.  
 DR PIR; A34401; A34401.  
 DR PIR; A34643; A34643.  
 DR PIR; A42038; A42038.  
 DR HSSP; P00794; 4CWS.  
 DR MEMOS; A01.010; -.  
 DR MIM; 116890; -.  
 DR InterPro; IPR001969; Asp-protease.  
 DR InterPro; IPR001461; Pepsin.  
 DR Pfam; PF00026; asp. 1.  
 DR PRINTS: PRO0792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 17  
 FT PROPEP 18 53  
 FT CHAIN 54 396  
 FT MOD\_RES 18 18  
 FT ACT\_SITE 96 96  
 FT ACT\_SITE 281 281  
 FT DISULFID 60 60  
 FT DISULFID 109 114  
 FT DISULFID 272 276  
 FT DISULFID 314 351  
 FT CARBOHYD 90 90  
 FT CARBOHYD 220 220  
 FT CARBOHYD 333 333  
 SO SEQUENCE 396 AA; 42793 MM; 40B643C5FB01521E CRC64;  
 Query Match 12.4%; Score 299.5; DB 1; Length 396;  
 Best Local Similarity 25.9%; Pred. No. 2.3e-17;  
 Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;



## EUKARYOTIC ASPARTYL PROTEASES FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: M59235; AA85369.1; -  
 DR PIR: C38302; C38302.  
 DR HSSP: P00790; IPSEN.  
 DR MEROPS: A01.001; -  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp; 1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolyase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Signal; Phosphorylation; Multigene family.  
 FT SIGNAL 1 15  
 FT PROPEP 16 59  
 FT CHAIN 60 387  
 FT MOD\_RES 129 129  
 FT ACT\_SITE 93 93  
 FT ACT\_SITE 276 276  
 FT DISULFID 106 111  
 FT DISULFID 267 271  
 FT DISULFID 310 343  
 SQ SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;

Query Match 12.2%; Score 296; DB 1; Length 387;  
 Best Local Similarity 26.9%; Pred. No. 4.3e-17;  
 Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

QY 30 YVEMTVGSPQPTLNILVDTGSSNFAVGAAPHF-----LHRYQROLSTYRLKRG 82  
 DB 75 YGCTISIGPQDFYVIFDTGSSNLMV---PSTYCSLACALHKRNPEDSSYQCTSEF 131  
 QY 83 VVVPYTGKWESEGLTDVLSIPHGPNVTVRANIAITESDKFFINSNMGGILGLAYAEI 142  
 DB 132 LSTYGTGSGMTGLGYDTKVGSIEDTNOIFGLSKTEPSITFLF--APPDGILGLAYPSI 189  
 QY 143 ARPDOSLEFPFSLVKQFHV-PULFSLQCGAGFPLNOSVLA SVGSGMIGIDSLVY 201  
 DB 190 SSSDAI--PVFDMMNMGILVSODLFVSIVLSSD-----EKSILVFGIDSYIT 237  
 QY 202 GSIWYPIRREMYEVIVRVEINGDLKM--DCKEYNTDKSIVDSGTNLRPKVFEA 259  
 DB 238 GSIWVPSVEGWMQITMDSVINGETIACADSC-----QAIVDTGSLTGP---TS 287  
 QY 260 AKSITAASTEFPPGFWLGEOLV-CMQAGTPNNIIPVISIYLMEGVTNOSFRITLIP 318  
 DB 288 AISNIQSYIGASR---NLIGENVISCSAIDSLPDIIV-----TING 325  
 QY 319 QOYLRFVEVDVANSQDDCYKFAISOSTGT-----VMGAVIMEGFVEVDRARRKIGFAV 372  
 DB 326 IQYPLPASAVIILKEDDCTSGLEGMMVDTYTGELMILGIVFIQYTFVDRANNOGLAA 385  
 QY 373 S 373  
 DB 386 A 386

RESULT 13  
 CATE\_CAVPO STANDARD; PRT; 391 AA.  
 ID CATE\_CAVPO  
 AC P25796;  
 DT 01-MAY-1992 (rel. 22, Created)  
 DT 01-MAY-1992 (rel. 22, Last sequence update)  
 DT 16-OCT-2001 (rel. 40, Last annotation update)

DE Cathepsin E precursor (EC 3.4.23.34).  
 GN CTSB.

OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathii; Cavidae; Cavia.  
 OX NCBI\_TaxID=10141;

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RN MEDLINE=92355614; PubMed=1644829;  
 RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koiwai O.,  
 RA Tanji M., Yakabe E., Athauda S.B., Takahashi K.;  
 RT "Gastric procathepsin E and progastricsin from guinea pig.  
 RT Purification, molecular cloning of cDNAs, and characterization of  
 RT enzymatic properties, with special reference to procathepsin E.";  
 RN J. Biol. Chem. 267:16450-16459(1992).

[2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Gastric mucosa;  
 RX MEDLINE=96073637; PubMed=8540321;  
 RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,  
 RA Tanji M., Athauda S.B., Takahashi K.;  
 RT "Isolation, characterization, and structure of procathepsin E and  
 RT cathepsin E from the gastric mucosa of guinea pig.";  
 RL Adv. Exp. Med. Biol. 362:211-221(1995).  
 CC -I- FUNCTION: DUE OF ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN  
 CC LYMPOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.  
 CC -I- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader  
 CC specificity.  
 CC -I- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: M88653; AA37052.1; -  
 DR EMBL: S80547; AA35844.1; -  
 DR PIR: A43356; A43356.  
 DR HSSP: P00794; 4CMS.  
 DR MEROPS: A01.010; -  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp; 1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 53  
 FT CHAIN 54 391  
 FT ACT\_SITE 92 92  
 FT ACT\_SITE 276 276  
 FT DISULFID 56 56  
 FT DISULFID 105 110  
 FT DISULFID 267 271  
 FT DISULFID 309 346  
 FT CARBOHYD 86 86  
 SQ SEQUENCE 391 AA; 42132 MW; 78D216BF8CFDABD CRC64;

Query Match 12.2%; Score 296; DB 1; Length 391;  
 Best Local Similarity 26.9%; Pred. No. 4.4e-17;  
 Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

QY 30 YVEMTVGSPQPTLNILVDTGSSNFAVGA---APHPFLHRYQROLSTYRLKRGVY 85  
 DB 74 YGCTISIGPQDFYVIFDTGSSNLMVPSVYCTSPACQTHPVHPSTISYREVGNFSFI 133  
 QY 86 PYTGKWESEGLTDVLSIPHGPNVTVRANIAITESDKFFINSNMGGILGLAYAEIAR 145

Db 134 QYNGSLTGTIGADQVSV-EGLTIVGQGFESVQEFKTFVH-AEFGILGLGYPSLAA-190  
146 DDLSEPFEDSLVOTHPNLFSLQLCAGFPLNQSEVLAVGGSMITGIDHSLYGSLSW 205  
191 -GGVTFPFDDMMKQ-----NVALPM-----FSVYSSNPGSGGSELFFGGYDPSHFGSLSN 241  
206 YTPIRREMYEVLIVVEINGODLKMCKEYNKDSIVDSGTNLRLPKVFEEAAKSTK 265  
242 WVPYTKQAVYQIALDLGIVG--DSVMFCSP--GCQAVIDGTSLITGP-----FGKITQLO 293  
266 AASTKFPDGFELGEOLYCWMQAGTPWNIPIVLSLYLNGEVTNOSFRI-----TILPOQ 320  
294 EALGATYVDEGY-----SVOC-----ANLNMMLDVT--FLINGVPTLNPFA 333  
QY 321 YLRPVEDVATSGDDCYKFAISOSTG-----TYMAVIMEGFYVVEDARRK 367  
Db 334 Y--TLIDPFDGMQVC-----STGFEGLEIQPPAGPLMILGDVPIROFYAVFEDRGNMR 383  
QY 368 IGFA 371  
Db 384 VGIA 387  
RESULT 14  
CATD\_HUMAN STANDARD; PRT; 412 AA.  
AC P07339;  
DT 01-APR-1988 (rel. 07, Created)  
DT 01-APR-1988 (rel. 07, Last sequence update)  
DT 01-MAR-2002 (rel. 41, Last annotation update)  
DE Cathepsin D precursor (EC 3.4.23.5).  
GN CTSD.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OY NCBI\_Taxid-9606;  
RX [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-85270436; PubMed-3927292;  
RA Faust P.L., Kornfeld S., Chirgwin J.M.;  
RT "Cloning and sequence analysis of cDNA for human cathepsin D.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87231068; PubMed-3588310;  
RA Westley B.R., May F.E.B.;  
RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive  
RL human breast cancer cells.";  
RN Nucleic Acids Res. 15:3773-3786(1987).  
RX [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91299158; PubMed-2069717;  
RA Redeker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;  
RT "Molecular organization of the human cathepsin D gene.";  
RL DNA Cell Biol. 10:423-431(1991).  
RN [4]  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE-94085791; PubMed-8262386;  
RA May F.E., Smith D.J., Westley B.R.;  
RT "The human cathepsin D-encoding gene is transcribed from an estrogen-  
RL regulated and a constitutive start point.";  
RN Gene 134:277-282(1993).  
RX [5]  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE-95021301; PubMed-7935485;  
RA Augereau P., Mitrallas F., Cavailles V., Gaudelot C., Parker M.,  
RA Rochefort H.;  
RT "Characterization of the proximal estrogen-responsive element of  
RL human cathepsin D gene.";  
RN Mol. Endocrinol. 8:693-703(1994).  
RX [6]  
RP SEQUENCE OF 170-180.

RC TISSUE=Liver;  
RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,  
RA Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,  
RA Appel R.D., Hughes G.J.;  
RN Submitted (JUN-1992) to the SWISS-PROT data bank.  
RP [7]  
RP VARIANT VAL-58  
RX MEDLINE-20179010; PubMed-10716266;  
RA Pappasoliotopoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,  
RA Maier W., Pauls J., Lautenschlager N., Heun R.;  
RT "A genetic variation of cathepsin D is a major risk factor for  
RT Alzheimer's disease.";  
RL Ann. Neurol. 47:399-403(2000).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).  
RC TISSUE=Spleen;  
RX MEDLINE-93223670; PubMed-8467789;  
RA Metcalf P., Fusek M.;  
RT "Two crystal structures for cathepsin D: the lysosomal targeting  
RL signal and active site.";  
RN EMBO J. 12:1293-1302(1993).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RC TISSUE=Liver;  
RX MEDLINE-93342076; PubMed-8393577;  
RA Baldwin E.T., Bhat T.N., Gulink S., Hosur M.V., Sowder R.C. II,  
RA Cachau R.E., Collins J., Silva A.M., Erickson J.W.;  
RT "Crystal structures of native and inhibited forms of human cathepsin  
RT D: implications for lysosomal targeting and drug design.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).  
CC -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR  
CC PROTEIN BREAKDOWN.  
CC -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,  
CC that of pepsin A. Does not cleave the 4-Gln-1-His-5 bond in B  
CC chain of insulin.  
CC -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- POLYMORPHISM: The Val-58 allele is significantly overrepresented  
CC in demented patients (11.8%) compared with nondemented controls  
CC (4.9%). Carriers of the Val-58 allele have a 3.1-fold increased  
CC risk for developing AD than noncarriers.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M11233; AAB59529.1; -  
DR EMBL; X05344; CAA28955.1; -  
DR EMBL; M63138; AAA51922.1; -  
DR EMBL; M63134; AAA51922.1; JOINED.  
DR EMBL; M63135; AAA51922.1; JOINED.  
DR EMBL; M63136; AAA51922.1; JOINED.  
DR EMBL; M63137; AAA51922.1; JOINED.  
DR EMBL; L12980; AAA16314.1; -  
DR EMBL; S74689; AAD14156.1; -  
DR EMBL; S52557; AAD13868.1; -  
DR PIR; A25771; KHHUD.  
DR PDB; 1LYA; 31-JAN-94.  
DR PDB; 1LYB; 31-JAN-94.  
DR MEROPS; A01.009; -  
DR SWISS-2DPAGE; P07339; HUMAN.  
DR Sienas-2DPAGE; P07339; -  
DR MTM; 116840; -  
DR InterPro; IPR001969; Asp-Protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp. 1.  
DR PRINTS; PR00792; PEPSIN.

KW	PROSTATE; PS00141; ASP. PROTEASE; 2.		
KM	Hydrolase; Aspartyl protease; Lysosome; Signal; Zymogen;		
KV	Polymorphisms; Alzheimer's disease; 3d-structure.		
FT	FT SIGNAL	1	18
FT	FT PROPEP	19	64
FT	FT CHAIN	65	412
FT	FT CHAIN	65	161
FT	FT CHAIN	169	412
FT	FT ACT_SITE	97	97
FT	FT ACT_SITE	295	295
FT	FT DISULFID	91	160
FT	FT DISULFID	110	117
FT	FT DISULFID	286	290
FT	FT DISULFID	329	366
FT	FT CARBOXYD	134	134
FT	FT CARBOXYD	263	263
FT	FT VARIANT	58	58
FT			
FT			
FT			
FT			
FT	STRAND	67	74
FT	FT TURN	75	77
FT	FT STRAND	78	85
FT	FT TURN	86	89
FT	FT STRAND	90	97
FT	FT TURN	98	99
FT	FT STRAND	103	107
FT	FT TURN	108	109
FT	FT TURN	112	113
FT	FT HELIX	115	118
FT	FT TURN	119	119
FT	FT STRAND	123	123
FT	FT HELIX	125	127
FT	FT TURN	129	130
FT	FT STRAND	132	141
FT	FT STRAND	146	158
FT	FT STRAND	172	184
FT	FT HELIX	188	192
FT	FT STRAND	197	200
FT	FT HELIX	204	206
FT	FT HELIX	208	210
FT	FT HELIX	214	220
FT	FT TURN	221	222
FT	FT STRAND	228	233
FT	FT STRAND	243	247
FT	FT TURN	248	248
FT	FT HELIX	252	254
FT	FT STRAND	255	263
FT	FT STRAND	267	267
FT	FT TURN	268	269
FT	FT STRAND	270	279
FT	FT TURN	280	281
FT	FT STRAND	284	285
FT	FT TURN	287	288
FT	FT STRAND	290	294
FT	FT TURN	296	297
FT	FT STRAND	298	298
FT	FT STRAND	301	303
FT	FT HELIX	305	315
FT	FT TURN	316	316
FT	FT STRAND	318	319
FT	FT TURN	322	323
FT	FT STRAND	325	328
FT	FT HELIX	329	334
FT	FT STRAND	338	342
FT	FT TURN	343	344
FT	FT STRAND	345	349
FT	FT HELIX	351	354
FT	FT STRAND	355	357
FT	FT TURN	359	362
FT	FT STRAND	365	368
FT	FT STRAND	370	372

FT	TURN	377	379
FT	TURN	383	385
FT	HELIX	387	390
FT	TURN	391	392
FT	STRAND	393	398
FT	TURN	399	402
FT	STRAND	403	409

Query Match 12.2%; Score 295.5; DB 1; Length 412;  
 Best Local Similarity 28.5%; Pred. No. 5,2e-17;  
 Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15

QY	30	YVEMTVGSPPTLNILVDTGSSNFAVGAHPFL-----HRYVRQLSSTYYRDLRKGV	83
QY <td>79</td> <td>YGEIGDIGPPQCFVTFVFDTGSSNLVPSIHCKLIDIAOMIHKKYMSDKSSTYYKNGTSE</td> <td>138</td>	79	YGEIGDIGPPQCFVTFVFDTGSSNLVPSIHCKLIDIAOMIHKKYMSDKSSTYYKNGTSE	138
QY <td>84</td> <td>YVPYTGKMEGELGNDLVIP-----HGPNVTVRANIAITFSDKFFINGSNWEGI</td> <td>134</td>	84	YVPYTGKMEGELGNDLVIP-----HGPNVTVRANIAITFSDKFFINGSNWEGI	134
DB <td>139</td> <td>DIHYGSGSLSGYLSQPTVSPCOSASSASALGCVKERYOVFGATQPGTTFLIAAFEDI</td> <td>198</td>	139	DIHYGSGSLSGYLSQPTVSPCOSASSASALGCVKERYOVFGATQPGTTFLIAAFEDI	198
QY <td>135</td> <td>LGALVAEIARPDLSLEPFEDSLVKQTHV-PNLESLDLCAGAPLNSOEVLYASVGSMTIG</td> <td>193</td>	135	LGALVAEIARPDLSLEPFEDSLVKQTHV-PNLESLDLCAGAPLNSOEVLYASVGSMTIG	193
DB <td>199</td> <td>LGMAVPRIS--VNNVLVPVDNLMQKLTVDONIRSFY-----SRDPDAQGCGELMLG</td> <td>248</td>	199	LGMAVPRIS--VNNVLVPVDNLMQKLTVDONIRSFY-----SRDPDAQGCGELMLG	248
QY <td>194</td> <td>GIDHSLYTSGSLMTPIPRREMYEYIVRPHI-NGOOLKMDCKEYNDKSIYDSGTINLRL</td> <td>252</td>	194	GIDHSLYTSGSLMTPIPRREMYEYIVRPHI-NGOOLKMDCKEYNDKSIYDSGTINLRL	252
DB <td>249</td> <td>GUDSKYKYGSLSLANTRKAYWGHLDQVEVASGLTL--CKE--GCEALVDYGTSLMVG</td> <td>303</td>	249	GUDSKYKYGSLSLANTRKAYWGHLDQVEVASGLTL--CKE--GCEALVDYGTSLMVG	303
QY <td>253</td> <td>PKVFEAAVYSIKAASTEFKPDGFWLGEOLV-CWQAGTTPWNIFFVYLSLYMGEYTNOS</td> <td>311</td>	253	PKVFEAAVYSIKAASTEFKPDGFWLGEOLV-CWQAGTTPWNIFFVYLSLYMGEYTNOS	311
DB <td>304</td> <td>PVDEVELDKAIGAVPLIQ-----GEVMIPCEKYST-----LPAITLKLGG---KG</td> <td>346</td>	304	PVDEVELDKAIGAVPLIQ-----GEVMIPCEKYST-----LPAITLKLGG---KG	346
QY <td>312</td> <td>FRITLIPQOYLPEVEDVANSQDCYFAISQ-----SSTGTWGAIVIMEGFYVVEDARK</td> <td>366</td>	312	FRITLIPQOYLPEVEDVANSQDCYFAISQ-----SSTGTWGAIVIMEGFYVVEDARK	366
DB <td>347</td> <td>YKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIIPPSGPLMLIGDVFGRITYVPRDNN</td> <td>402</td>	347	YKLS--PEDYTLKVSQAGKTL--CLSGFMGMDIIPPSGPLMLIGDVFGRITYVPRDNN	402
QY <td>367</td> <td>RIGFAVSA 374</td> <td></td>	367	RIGFAVSA 374	
DB <td>403</td> <td>RGVFAEAA 410</td> <td></td>	403	RGVFAEAA 410	

RESULT 15  
 PEP4-RABIT  
 ID PEP4.RABIT STANDARD; PRT; 387 AA.  
 AC P28713;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, last annotation update)  
 DE Pepsin II-4 precursor (EC 3.4.23.1) (Pepsin A).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Gnathata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxId=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91009127; PubMed=2129536;  
 RA Kageyama T., Tanabe K., Koiwai O.;  
 RT "Structure and development of rabbit pepsinogens. Stage-specific  
 RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and  
 RT gene expression during development.";  
 RL J. Biol. Chem. 265:17031-17038(1990).  
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE  
 CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.  
 CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
 CC HORMONES AND RELATED SUBSTANCES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC D38302.  
 DR PIR; D38302; D38302.  
 DR HSSP; P00790; IPBN.







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:36 ; Search time 32.31 seconds

(without alignments)  
1356.135 Million cell updates/sec

Title: US-09-724-571-43

Perfect score: 2419  
Sequence: 1 ETDEPEEPGRGRGSFVEMVD.....CLRLRQHDHDFADISLTK 456

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*

1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2419	100.0	501	2	A59090 aspartic proteinase
2	308	12.7	384	2	JC7574 pepsinogen A - Afr
3	308	12.7	385	2	JC7575 pepsinogen A - bul
4	306	12.6	387	2	B38302 pepsinogen C - Afr
5	303	12.5	383	2	JC7573 pepsinogen C - Afr
6	302.5	12.5	388	1	S19682 pepsin A (EC 3.4.2
7	301	12.4	382	1	PECH pepsin A (EC 3.4.2
8	299.5	12.4	396	2	A34401 pepsin A (EC 3.4.2
9	298.5	12.3	383	2	A41443 pepsin E (EC 3.
10	298.5	12.3	384	2	A39314 pepsin (EC 3.4.23.
11	296	12.2	387	2	C38302 pepsin (EC 3.4.23.
12	296	12.2	391	2	A43356 pepsin E (EC 3.
13	295.5	12.2	412	1	KHHUD pepsin E (EC 3.
14	295	12.2	387	2	D38302 pepsin D (EC 3.
15	289.5	12.0	388	1	S19684 pepsin (EC 3.4.2
16	287.5	11.9	444	1	T24204 hypothetical prote
17	287	11.9	407	1	KHRPD cathepsin D (EC 3.
18	286	11.8	387	2	E38302 pepsin (EC 3.4.23.
19	285.5	11.8	398	2	S66455 cathepsin E (EC 3.
20	285	11.8	398	1	I51185 cathepsin D (EC 3.
21	283.5	11.7	388	1	PEMOCAR pepsin A (EC 3.4.2
22	282.5	11.7	388	1	PEHU pepsin A (EC 3.4.2
23	281.5	11.6	381	1	CMSHR pepsin A (EC 3.4.2
24	281.5	11.6	388	1	PEMOAJ pepsin A (EC 3.4.2
25	281.5	11.6	388	1	A50142 pepsin A (EC 3.4.2
26	279.5	11.6	388	2	B30142 pepsin A (EC 3.4.2
27	279.5	11.6	410	1	KHMSD cathepsin D (EC 3.
28	278.5	11.5	386	1	PEPG pepsin A (EC 3.4.2
29	278	11.5	387	2	JC7245 pepsinogen A - com

30	277.5	11.5	380	2	I47176 chymosin (EC 3.4.2
31	277.5	11.5	396	2	S36865 cathepsin E (EC 3.
32	276	11.4	389	2	JEO371 pepsin C (EC 3.4.2
33	273.5	11.3	381	1	CMBO pepsin C (EC 3.4.2
34	270.5	11.2	377	1	PEMOJ pepsin (EC 3.4.2
35	270.5	11.2	389	2	A38302 pepsin (EC 3.4.23.
36	270	11.2	376	2	I45856 aspartic proteinase
37	268.5	11.1	344	1	KHPGD cathepsin D (EC 3.
38	267.5	11.1	381	2	JC7247 prochymosin - comm
39	266	11.0	380	2	S03433 candidapepsin (EC
40	266	11.0	405	2	A25379 saccharopepsin (EC
41	264	10.9	396	2	T47207 aspartic proteinase
42	263.5	10.9	388	2	JC7246 pepsinogen C - com
43	262.5	10.9	394	2	B43356 pepsinogen C (EC 3.4
44	261.5	10.8	387	2	A45117 aspartic proteinase
45	261.5	10.8	388	2	A29937 gastricsin (EC 3.4

#### ALIGNMENTS

##### RESULT 1

A59090 aspartic proteinase (EC 3.4.23.-) BACE precursor - human

N:Alternate names: beta-secretase; beta-site APP cleaving enzyme

C:Species: Homo sapiens (man)

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-May-2000

C:Accession: A59090

R:Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Teplio M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro Science 286, 735-741, 1999

A>Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran

A:Reference number: A59090; MUID:20002972

A>Note: submitted to Genbank, September 1999

A:Accession: A59090

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-501 <VAS>

A:Cross-references: GB:AF190725; NID:G6118538; PID:NARF04142.1; PID:G6118539

C:Genetics:

A:Gene: BACE

C:Superfamily: beta-secretase

C:Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-45/Domain: propeptide #status predicted <PRO>

F:46-501/Product: acid proteinase BACE #status predicted <MAT>

F:461-477/Domain: transmembrane #status predicted <TRN>

F:93,289/Active site: Asp #status predicted

F:153,172,223,354/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2419; DB 2; Length 501;

Best Local Similarity 100.0%; Pred. No. 3.9e-192;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ETDEPEEPGRGRGSFVEMVDNLGRSGGGYVEMVGSPPQTLNLTVDGSSNFVAGAP	60
DB	46	ETDEPEEPGRGRGSFVEMVDNLGRSGGGYVEMVGSPPQTLNLTVDGSSNFVAGAP	105
QY	61	HPFLHRYQRLSLSTYRDLRKGYYVYTGKWEGLGTDLSIPHGPNVYVRANATATTE	120
DB	106	HPFLHRYQRLSLSTYRDLRKGYYVYTGKWEGLGTDLSIPHGPNVYVRANATATTE	165
QY	121	SDKEPFGNSNNEGILGLAVAEIARDDSLPEPDSLVQTHVPLNFSIQLCGAGPPLNOS	180
DB	166	SDKEPFGNSNNEGILGLAVAEIARDDSLPEPDSLVQTHVPLNFSIQLCGAGPPLNOS	225
QY	181	EVLASVGGSMIIIGCIDHSLYTGSLWYTPIRREMYEVIIVREINIGODLKMDCREYNDK	240
DB	226	EVLASVGGSMIIIGCIDHSLYTGSLWYTPIRREMYEVIIVREINIGODLKMDCREYNDK	285
QY	241	SIVDSGTTNLRPLPKKVFPAVAKSIKAAASSTEFDPDGFWLGQLYCWAQGTTPWNIFPVIS	300

```
|||||
Db 286 STVDSGTNLRPKVFFRAVAKSIKAASSTKEFPDGFHLEGLVCMQAGTTPWNIFFPVIS 345
QY 301 LYLMGEVNTQSFRTITLPOQYLRPEVDVATSDDCYKFAISQSSGTGYWGAIVMEGFYV 360
Db 346 LYLMGEVNTQSFRTITLPOQYLRPEVDVATSDDCYKFAISQSSGTGYWGAIVMEGFYV 405
QY 361 FDRARRRIGFAVNSACHVDEFTFAVEGPFVTLDMEDCGYNIPTDESTLMTIAYVMAI 420
Db 406 FDRARRRIGFAVNSACHVDEFTFAVEGPFVTLDMEDCGYNIPTDESTLMTIAYVMAI 465
QY 421 CALFMPLPLCMWCQWRCLRCLRQHDHDFADISLKL 456
Db 466 CALFMPLPLCMWCQWRCLRCLRQHDHDFADISLKL 501

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7574; PC7119
R.Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Molecule type: protein
A:Residues: 16-35;57-76 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a
A:Gene: Pga
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 12.7%; Score 308; DB 2; Length 384;
Best Local Similarity 27.5%; Pred. No. 1.3e-17;
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;

Db 72 YVETISIGTPPEFTVIFDTSANLWV--PSVYSSQACSNHNRNPQSSSTPQATNTP 128
QY 30 YVEMTVGSPPTQNLIVDTGSSNFAVGAAPHPFL-----HRYQRQLSTYRDLRKG 82
Db 72 YVETISIGTPPEFTVIFDTSANLWV--PSVYSSQACSNHNRNPQSSSTPQATNTP 128
QY 83 VVVPYTOGKEGELGTDLVSIPHGPNVTVRANIAITESDK-FFINGSNMEGILGLAVAE 141
Db 129 VSIQOTGSMGFLGVDITQV---GNIQISNMGFLSEBPSFLYSPFDGILGLAPFS 185
QY 142 IARPDDSLPEPFDLVKQTHVP-NLFSLOLCGAGPPLNQSEVLASVSGSMITIGDHSLSL 200
Db 186 IA--SSQATPVPDNNMWSQGLIPQNLFSVYLSSPG-----QTSVYLFEGVDNSYY 233
QY 201 TCSLWVTPLRREMYEVIIIVRVEINGQDL--KMDCKEYNDKSIYVSGTTLRLPKVEE 258
Db 234 SGLSNVPLTAETIYQITLDSVINSQVACSSQSC-----QAIYDTGSLMGPSTPI- 286
QY 259 AAVKSIKAASTKEFPDGFHLEGLV-CMQAGTTPWNIFFPVISLYLMGEVNTQSFRTITL 317
Db 287 ANIQNYIGASQSSN-----GQYVINCNNISMPTIVF-----TTN 321
QY 318 POQY-LRPEVDVATSDDCYK-FAISQSSGT-----VMGAIVMEGFYVFDARRKRIGFA 371
Db 322 GVOYPLSPAYVRQNOGSSGFGQAMNLPITNSGDLWITLGVFIROYFTVFDRANNYVAIA 381

RESULT 3
JC7575
pepsinogen A - bullfrog
```

```
C:Species: Rana catesbeiana (bullfrog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7575
R.Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinog
A:Reference number: JC7573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: JC7575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-lik
C:Gene: Pga
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 12.7%; Score 308; DB 2; Length 385;
Best Local Similarity 27.8%; Pred. No. 1.3e-17;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;

QY 30 YVEMTVGSPPTQNLIVDTGSSNFAVG---AAPHPFLHRYQRQLSTYRDLRKGYYV 85
Db 73 YVETISIGTPPEFTVIFDTSANLWVSVYSSQACSNHNRNPQSSSTPQATNTPVSI 132
QY 86 PYTOGKWEGELGTDLVSIPHGPNVTVRANIAITESDK-FFINGSNMEGILGLAVAEIAR 144
Db 133 QYGTGSMGFLGVDITQV---GNIQITNOIFGLSSEPSFLYSPFDGILGLAPSLA- 188
QY 145 PDDSLPEPFDLVKQTHVP-NLFSLOLCGAGPPLNQSEVLASVSGSMITIGDHSLYTGS 203
Db 189 -SSQATPVPDNNMWSQGLIPQDLFSVYLSQG-----GSFVLEGGVDITYTGN 237
QY 204 LMYTPLRREMYEVIIIVRVEINGQDLK--DCKEYNDKSIYDSTTLRLPKVEEAV 261
Db 238 LNVVPLTAETIYQITVDSISIGQVYACSGSC-----SAIVDTGSLLAGF--STPI 287
QY 262 KSIKAASTKEFPDGFHLEGLVCMQAGTTPWNIFFPVISLYLMGEVNTQSFRTITLPOQY 321
Db 288 ANIQYIGANDQNSQGY---INCNNISMPTIVF-----TINGVQY 326
QY 322 LRPEVD-VATSDDC--YKFAISQSSGT--VMGAIVMEGFYVFDARRKRIGFA 371
Db 327 PLPASAYVRQNOGSSGFGQAMNLPITNSGDLWITLGVFIROYFTVFDRANNYVAMA 382
```

```
RESULT 4
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R.Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.6%; Score 306; DB 2; Length 387;
Best Local Similarity 27.1%; Pred. No. 2e-17;
Matches 98; Conservative 67; Mismatches 131; Indels 66; Gaps 15;

QY 30 YVEMTVGSPPTQNLIVDTGSSNFAVG---AAPHPFLHRYQRQLSTYRDLRKGYYV 85
Db 327 PLPASAYVRQNOGSSGFGQAMNLPITNSGDLWITLGVFIROYFTVFDRANNYVAMA 382
```

Db 75 YFCTISCTPPEFTYIPTGSSNLNWPSTYSSSLACFLHKRPDDSTFQATSETLSI 134  
 QY 86 PYIQGMESELGTDLVSIPIHPNVTVRANIAAITESD---KEPINSNMEGILGLAYAEI 142  
 Db 135 TYGTGSMTGILGIDYKVA---GNIEDTNOIFGLSKTEPGITFLV--APPDGILGLAYPSI 189  
 QY 143 ARPDDELPEFPDVLVQOTYV--PNLFSLOLCGAGFPLNOSEVLASVGSMTIGIDHSLYT 201  
 Db 190 SASDAT--PVFDNMNMEGLVSEDLFSVYLSNMG-----EKSMVMEFGIDSSYTT 237  
 QY 202 GSLMYPIRREMYEVIYIRVEINSGDLKM--DCKEYNKDSIVDSGTNNLPRKVFEEA 259  
 Db 238 GSLMYPIRREMYEVIYIRVEINSGDLKM--DCKEYNKDSIVDSGTNNLPRKVFEEA 259  
 QY 260 AVKSIKAASSTKFPDGFMLGEOLV-CMOAGTTPMNIFFVISLYLMGEVTNOSFRITILP 318  
 Db 292 IQSYIGASKNL-----LGENITSCSAIDSLPDIVF-----TINN 325  
 QY 319 QOYLRFVED-VATSDDC---YKFAISOSTGT--VMGAVIMEGFYVFDRAKRIGFAV 372  
 Db 326 VQYPLPASAVYLIKEDDCLSGFDGMNLDTSYGEMLTGDFVFIROYFTVDRANNOVGLAA 385  
 QY 373 SA 374  
 Db 386 AA 387

# RESULT 5

JCT573  
 A:Accession: JCT573  
 A:Molecule type: protein  
 A:Residues: 1-383 <IKU>  
 A:Cross-references: DDBJ:AB045379  
 A:Accession: PC7118  
 A:Molecule type: protein  
 A:Residues: 1-768 <IKU>  
 C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like a  
 C:Gene: PGC  
 A:Superfamily: pepsin  
 C:Keywords: stomach; zymogen

Query Match 12.5%; Score 303; DB 2; Length 383;  
 Best Local Similarity 27.4%; Pred. No. 3,4e-17;  
 Matches 104; Conservative 57; Mismatches 122; Indels 96; Gaps 17;

QY 30 YVENVTVSPQOTLNLIVDTGSSNFANVGA-----APHFLHRYXOROLSTSYRDLRK 81  
 Db 67 YVGEISITGPONLNLIVDTGSSNLNWPSTYSSSLACFLHKRPDDSTFQATSETLSI 134  
 QY 82 GVVYVYTGKMEGELGTDLVSIPIHPNVTVRANIAAITESD---KEPINSNMEGILGLAYAEI 142  
 Db 123 QFSLQYIGASKNL-----LGENITSCSAIDSLPDIVF-----TINN 325  
 QY 137 LAYEIAIAPDDELPEFPDVLVQOTYV--PNLFSLOLCGAGFPLNOSEVLASVGSMTIGIDHSLYT 201  
 Db 176 LAYPSIA--VCGATTVMQGMQO---NLNQPI--FGFYLSSGS--SQNGEVAIFGVD 225  
 QY 197 HSLTGSIMTYPIRREMYEVIYIRVEINSGDLKM--DCKEYNKDSIVDSGTNNLPRKVFEEA 259  
 Db 226 ONYTTGQIYMPVSEYTWQIGGFSINGOATGWCSCOG-----QATVDGTGTSLLTAP 279

QY 254 KVFEEAVKSIKAASSTKFPDGFMLGEOLVCMOAGTTPMNI--FPVISLYLMG----- 305  
 Db 280 QVYFSSLSIISIAQODN-----GQYVSCS-----NQNLPITISFTISGVSPFLP 325  
 QY 306 ---EVTNDS-----FRITLPOQYLRPEVDVATSDCKYKFAISOSTGTVMGAVIME 355  
 Db 326 PAAVYLQSSGCTIGIMPTLYPSQNGPL-----WILGDVFLR 364  
 QY 356 GFYVVEDRAKRIGFAVSA 374  
 Db 365 EYYSVYDLGNNOVGATTA 383

# RESULT 6

S19682  
 A:Accession: S19682  
 A:Molecule type: mRNA  
 A:Cross-references: EMBL:X59753; NID:938070; PIDN:CAA42425.1; PID:938071  
 A:Note: parts of sequence, including amino ends of pepsinogen and activation intermedi  
 C:Comment: This is a minor component of pepsin at all post-partum stages.  
 C:Superfamily: pepsin  
 C:Keywords: aspartic proteinase; gastric juice; hydrolyase; phosphoprotein; protein di  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-38/Domain: activation peptide #status experimental <APR>  
 F:63-38/Domain: product: pepsin A 4 #status experimental <ENZ>  
 F:38-39/Cleavage site: Leu-Lys (pepsin) #status experimental  
 F:62-63/Cleavage site: Asp #status predicted  
 F:94,27/Active site: Asp #status predicted  
 F:107-112,268-272,311-344/Disulfide bonds: #status predicted  
 F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 12.5%; Score 302.5; DB 1; Length 388;  
 Best Local Similarity 27.6%; Pred. No. 3,8e-17;  
 Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;

QY 3 DEEPPEGRSGFEVMDNLNKGSGQGYVEMTVGSPQOTLNLIVDTGSSNFANVGAAPHP 62  
 Db 64 DEQPLE-----NYLDV-----EYFGTIGCTPAQNTFTVFDGSSNLNWP---PSV 105  
 QY 63 FL-----HRYXOROLSTSYRDLRKGYVYPTGKMEGELGTDLVSIPIHPNVTVRANI 115  
 Db 106 YCYSLACMDHNLFPDOSTSYRATSKVSYTYGSMTGILGIDYKVA---GGISDTNQI 162  
 QY 116 AAITESDK--FINSNMEGILGLAYAEIARPDDELPEFPDVLVQOTYV--PNLFSLOLCGA 173  
 Db 163 FGLSETPEGFLLYARPDGILGLAYPSI---SSGATYVFDIMNQRLVSDLSVYLSAD 220  
 QY 174 GFPLNOSEVLASVGSMTIGIDHSLYTGLSWTYPIRREMYEVIYIRVEINSGDL--KM 231  
 Db 221 ---DQS-----GSVYIFEGIDSSYTTGSLNWPVSEGEVWQISVDSITMNGTITCAK 270  
 QY 232 DCKEYNKDSIVDSGTNNLPRKVFEEAVKSIKAASSTKFPDGFMLGEOLV-CMOAGT 259  
 Db 271 GC-----QATVDGTSILGTPSPNINIOSDIGASNSD-----GEMVYSCSAISS 316  
 QY 291 TPWNIFFVISLYLMGEVTNOSFRITILPOY--LRPVEDVATSDDCYK-----FAISOSS 344  
 Db 317 LPLDIV-----TINGVOYPLPSPAVIILQSGSGCTSGRGQADVPTESG 358  
 QY 345 TGYVAGAVIMEGFYVFDRAKRIGFA 371



```

Db      170 TEPOQTETVD-AEFDGIIIGLGYPSLA--VGVTPVFDNMMAQ----NLVDPMPMSVYSSN 222
QY      179 QSEVLASVSGSMITGIDHSLYTGSMTWTPLRREMYEVLIVREINQGLKMDCKEYNT 238
Db      223 PE---GGAGSGLIFGGYDHSFSGSLMWVPVTAQVWQIALDNIQVGG--TVMECSF--G 275
QY      239 DKSTVDSGTTNLRLPKRVFAAVSKRAASSTKEKFPDGFMLGRLVCWQAGTTPWNTFPV 298
Db      276 COAIVDTGTSLITGSPDKIKOLONAICAP-----VDGEFAVE-----CANLVMPD 322
QY      299 ISLYLMEVYNQSFRTITLQQYLKRPVEDVATSODDCYKFAISQSTG----- 346
Db      323 VTFTTNG-----VPRYLSPTAY--TLDPVDMQFC-----SSGFGGLDIHPAPG 366
QY      347 -TVAGAVIMEGFYVFPDRARRRIGFA 371
Db      367 LMTLGDVFIRQFYSVFDRGNRRVGLA 392

```

```

RESULT 9
A41443
pepsin (EC 3.4.23.-) precursor, embryonic - chicken
C:Species: Gallus gallus (chicken)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 21-Jul-2000
C:Accession: A41443
R:Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.
J. Biochem. 103, 290-296, 1988
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsin
A:Reference number: A41443; MUID:86227903
A:Accession: A41443
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-383 <NAY>
A:Cross-references: GB:D00215; NID:92760810; PIDN:BA00153.1; PID:g222853
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; protein digestion

```

```

Query Match      12.3%; Score 298.5; DB 2; Length 383;
Best Local Similarity 25.2%; Pred. No. 8e-17;
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY      30 YVEMTVGSPQPTLNLIVDTGSSNFAVGA---APHPFLHRYQQLSSTYDLEKRYIV 85
Db      76 YGTISTIGTPDQFTYVFTGSSNLMWVPVSCTSPACSHQMFNSQSTYKSTGQNTSI 135
QY      86 PYTQGMKEGLGTDLVSIPLGPNVTVRANIAITSDKFFINGSMWEGILGLAYEAIARP 145
Db      136 HGTGDMECTVGCDFVTVASLMDTNQLFGLST-SFPGQFFVY-VKFDGILGLGYSLAH- 192
QY      146 DDLSEFPDLSVKQTHV-PNLFSLDLCGAGPLNQSEVLASVSGSMITGIDHSLYTGS 204
Db      193 -GGITPVEDNMVNESILDEQNLFSVYLS-----REPMSVAVPGIDESYFTGSI 240
QY      205 WYTPLRREMYEVLIVREINQGLD--KMDCKEYNDKSYNDSGTNLRPKRYFEAAVK 262
Db      241 NMIPSYGYQIMQISMDSTIVNQELTACSSG-----QALIDTGSTLVAGAPASINDIOS 294
QY      263 SIKAASTKEKFPDGFMLGEQLVCWQAGTTPWNTFPVLSL---YLMGEVYNQSFRTITL 318
Db      295 AVGANQNT-----YGEYSV-----NCSHILAMPVAVFVIGI----- 326
QY      319 OQYLKRPVEDVA---TSQDDCKKFAISQSSGTGVAGVIMEGFYVFPDRARRRIGFA 371
Db      327 -QY--PVPAALATTEONGGQTCMSSFQNSSADLMTLIGDVFIRYVYSIFDRANNRVGLA 380

```

```

RESULT 10
A39314
gastricsin (EC 3.4.23.3) precursor - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 22-Jun-1999

```

```

C:Accession: A39314
R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kag
J. Biol. Chem. 266, 22436-22443, 1991
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and
A:Reference number: A39314; MUID:92042186
A:Accession: A39314
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-384 <YAK>
A:Cross-references: GB:M73750; NID:9213687; PIDN:AAA49530.1; PID:g213688
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; protein digestion

```

```

Query Match      12.3%; Score 298.5; DB 2; Length 384;
Best Local Similarity 25.5%; Pred. No. 8.1e-17;
Matches 100; Conservative 61; Mismatches 142; Indels 89; Gaps 15;

```

```

QY      14 SFVEMVDNLKSGSGGYVEMTVGSPQPTLNLIVDTGSSNFAVGAAPHPFL-----HR 66
Db      51 NFPAIFEEPLANTMDMSYGEISIGTPQNFVLFPDTGSSNLMV--PSTYCQSQACTNHP 107
QY      67 YQROLSTYTPDLRKGVVVPYQGMKEGLGTDLSIPHGPNVTVRANIA-----AI 118
Db      108 QPNPQSSSYSSNQQPFSLQYCTGSLTGLIGDYQI-----QNTAISQDEFGLSV 158
QY      119 TSDKFFINGSMWEGILGLAYEAIARPDSLEPPFDSLVKQTHVPN-LFSQLCGAGFPL 177
Db      159 TEPGTNFPYV-AQFDGIIILAVPESIA--EGGATTVWQMGIQNLNLQPLFARYLSGQGNQSQ 215
QY      178 NOSEVLASVSGSMITGIDHSLYTGSMTWTPLRREMYEVLIVREINQGLD---LKMDCK 234
Db      216 N-----GGEVAVGVDQNTNLSQILTWPTVSTETWQIDIGFSVNGATGMC SQGC- 266
QY      225 EYVNDKSYDVSQTTNLRPKRVFEAAVSKRAASSTKEKFPDGFMLGEQLV-CWQAGTTPW 293
Db      267 -----QGIVDTGTSILTFAPQSVFSSLMQSIGAQODQN-----GQVAVCSNIGLSIPT 313
QY      294 NIFPVI-----SLYLMGEVYNQ---SFRITTLPOQLKRPVEDVATSODDCYKFAISQ 342
Db      314 ISFTISGVSPPLPSAVYVLAQNSGYCTGTGIMPTYLPSPONGPL----- 356
QY      343 SSTGTWGAIVMEGFYVFPDRARRRIGFAVSA 374
Db      357 ----WIIGDVFRLQRYYSYVDLGNNOVGFAAANA 384

```

```

RESULT 11
C38302
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997
C:Accession: C38302
R:Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: C38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

```

```

Query Match      12.2%; Score 296; DB 2; Length 387;
Best Local Similarity 26.9%; Pred. No. 1.3e-16;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

QY      30 YVEMTVGSPQPTLNLIVDTGSSNFAVGAAPHPF-----LRRYQROLSTYTPDLRK 82
Db      75 YFGISTIGTPDQFTYVFTGSSNLMV--PSTYCSSLACALHKKRNPEDSSTYQGTSET 131

```





RESULT 14  
D38302  
pepsin (EC 3.4.23.-) II-4 precursor - rabbit  
C:Species: *Oryctolagus cuniculus* (domestic rabbit)  
C:Date: 14-Jun-1991 #sequence\_rev1sion 20-Sep-1991 #text\_rev1sion 23-Feb-1997  
C:Accession: D38302  
R:Kageyama, T.; Tanabe, K.; Kotwal, O.  
J. Biol. Chem. 265, 17031-17038, 1990  
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucle  
A:Reference number: A38302; MUID:91009127  
A:Accession: D38302  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-387 <KAG>  
A:Cross-references: GB:M59235; GB:J05638  
C:Superfamily: pepsin  
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.2%; Score 295; DB 2; Length 387;  
Best Local Similarity 26.1%; pred. NO. 1.6e-16;  
Matches 97; Conservative 65; Mismatches 123; Indels 86; Gaps 14

C: Superfamily: pepsin  
C: Activation: two step activation is observed, activation is predominantly a  
C: Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di  
F:1-15/Domain: signal sequence #status predicted <Sig>  
F:16-388/Product: pepsinogen A 2/3 #status experimental <Ppt>  
F:16-63/Domain: activation peptide #status experimental <Apt>  
F:63-388/Product: pepsin A 2/3 #status experimental <ENZ>  
F:40-41/Cleavage site: Asp-Phe (pepsin) #status experimental  
F:62-63/Cleavage site: Leu-Ile (pepsin) #status experimental  
F:94,277/Active site: Asp #status predicted  
F:107-112,268-272,311-344/Disulfide bonds: #status predicted  
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match	12.0%;	Score 289.5;	DB 1;	Length 388;
Best Local Similarity	27.0%;	Pred. No. 4.6e-16;		
Matches 103; Conservative	66;	Mismatches 141;	Indels 71;	Gaps 16.

QY	3	DEPEEPRGRSEFVEMVDNLGRKSGGGYVYKEMGSPPTLAILVDYTGSSNFAYGAAPHP	62
Db	64	DEQPLE-----NYIDM-----EYETIGTIGTPADFTVITDTSNLMWPSSYCS	108
QY	63	FL-----HHYVROLSSYYRDLRKGYVYPTQCKMBEGLCTDLYVIPGPNVTYANTAAI	118
Db	109	SLACNNHRRFPQSSSTVYSTGCTSVITYGTSAGMGLGAYDVQV-----GGISDINQIFGL	165

```

OY 119 TESDK-FEINGSNMEGILGLAYAEIARPPDLSLEPFDSLVRKQTHV-PNLFSLQICGAGFP 176
Db 166 SETEPGSEFLIYARPDGILGLAYPSTIS--SSGAPFVFDNIMWQGLVSODLFSVYLSAD--- 220
OY 177 LNQSEVLASVGGSMITIGIDHSLYTGSLMTPIREMYEYIIVRVEINQODLKMCKEY 236
Db 221 -DQS-----GSVVIFGIDSSYTYTGLMWVPVSEGVQISVDSTIMNGEALA--CAE- 270
OY 237 NYDKSIYDSCGTNTLRPKKVEEAALKSIKASTKEKFPDGFGLGEOLV-CWQAGTTPWNI 295
Db 271 -GCGRIYDGTSLTGTPTSPITANTIOSDIGASEND-----GEMVWSCSAISSLPDIY 321
OY 296 PPVISTLYMGEVITNQSFRITLLPOQYLRPEVDVATSQDDCYK-----FAISQSTGYMG 350
Db 322 F-----TINGIQYPPPSAY-----ILOSQSGCISGFQGMQDVPTESGELWITLG 364
OY 351 AVIMEGFYVFPDRAKRRTGFA 371
Db 365 DVEIRQYFTVEDRANNOVOLA 385

```

Search completed: August 7, 2002, 09:14:17  
Job time: 41 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:35 ; Search time 22.58 seconds

(without alignments)  
493.272 Million cell updates/sec

Title: US-09-724-571-43

Perfect score: 2419

Sequence: 1 ETDEPEEPGRSGFVEMVD.....CLRLRQHQHDFADISLKLK 456

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCrus.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2413	99.8	501	4	US-09-009-191-2
2	2320	95.9	774	4	US-09-009-191-4
3	1171.5	48.4	518	3	US-08-999-723-2
4	1171.5	48.4	518	4	US-09-434-427-2
5	1146.5	47.4	514	4	US-09-717-432-2
6	1146.5	47.4	514	4	US-09-912-484-2
7	299.5	12.4	396	1	US-08-208-007A-13
8	299.5	12.4	396	4	US-09-032-523-9
9	295.5	12.2	412	1	US-08-208-007A-12
10	295.5	12.2	412	4	US-08-974-691-4
11	279.5	11.6	458	6	5217891-15
12	273.5	11.3	409	1	US-09-640-305-6
13	273.5	11.3	409	1	US-08-360-673-6
14	273.5	11.3	427	2	US-08-846-021A-8
15	270	11.2	410	1	US-08-088-633-2
16	270	11.2	410	1	US-08-245-756-2
17	270	11.2	410	1	US-08-441-750-2
18	270	11.2	410	2	US-08-441-751-2
19	270	11.2	410	5	PCT-US92-02521-2
20	250	10.3	349	4	US-09-032-523-3
21	227	9.4	398	1	US-08-328-314-2
22	227	9.4	398	4	US-08-731-045-2
23	212	8.8	419	4	US-08-974-691-3
24	210	8.7	397	3	US-09-079-415-2
25	196.5	8.1	430	1	US-08-535-237-2
26	195.5	8.1	427	1	US-07-958-222A-2
27	193	8.0	330	3	US-08-115-753-1

28	193	8.0	419	3	US-08-115-753-2	Sequence 2, Appl1
29	193	8.0	419	3	US-08-115-753-3	Sequence 33, Appl1
30	187	7.7	420	4	US-09-008-271A-4	Sequence 4, Appl1
31	187	7.7	420	4	US-08-974-691-8	Sequence 8, Appl1
32	184.5	7.6	395	1	US-08-723-938-3	Sequence 3, Appl1
33	184.5	7.6	395	1	US-09-080-538-3	Sequence 3, Appl1
34	184	7.6	445	4	US-08-974-691-6	Sequence 6, Appl1
35	184	7.6	451	4	US-08-974-691-2	Sequence 2, Appl1
36	149	6.2	437	4	US-09-353-332-2	Sequence 2, Appl1
37	128.5	5.4	140	3	US-09-211-631-13	Sequence 13, Appl1
38	128.5	5.4	140	4	US-09-265-628-13	Sequence 13, Appl1
39	129.5	5.4	140	4	US-09-001-141-11	Sequence 11, Appl1
40	129.5	5.4	140	4	US-09-532-803-6	Sequence 6, Appl1
41	129.5	5.4	140	4	US-09-653-403-14	Sequence 14, Appl1
42	97	4.0	1030	4	US-09-091-117-2	Sequence 2, Appl1
43	95.5	3.9	280	4	US-09-160-246-14	Sequence 2, Appl1
44	92.5	3.8	1097	2	US-08-680-326-39	Sequence 39, Appl1
45	88	3.6	746	2	US-08-838-219B-6	Sequence 6, Appl1

## ALIGNMENTS

RESULT 1  
Sequence 2, Application US/09009191  
Patent No. 6319689  
GENERAL INFORMATION:  
APPLICANT: POWELL, DAVID  
APPLICANT: CHAPMAN, CONRAD  
APPLICANT: MURPHY, KAY  
APPLICANT: SMITH, TRUDI  
TITLE OF INVENTION: ASP2  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESS: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,191  
FILING DATE: 20-JAN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: UK 9701684.4  
FILING DATE: 28-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70368  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-009-191-2  
Query Match 99.8%; Score 2413; DB 4; Length 501;  
Best Local Similarity 99.8%; Pred. No. 3.3e-251;

	Matches	455, Conservative	0, Mismatches	1, Indels	0, Gaps	0, Nucleotide
QY	1	ETDEEPEEBGRRGSTVEVMDNLKRGSGGYIYEMTVGSPPTLLNLVDTGSSNFAVGAP	60			
Db	46	ETDEEPEEBGRRGSTVEVMDNLKRGSGGYIYEMTVGSPPTLLNLVDTGSSNFAVGAP	105			
QY	61	HPFLRRYYOROLSSYYRDLRKGVVYPYTGKKEGELGTGLVSIPIHPNPTVANAIAITE	120			
Db	106	HPFLRRYYOROLSSYYRDLRKGVYEPYTGKKEGELGTDLVSIPIHPNPTVANAIAITE	165			
QY	121	SDKFEINGSNMEGILGLAYAEIARPDSDLPEPDSLVKOTHPNPLESLDLCAGAPLNOS	180			
Db	166	SDKFEINGSNMEGILGLAYAEIARPDSDLPEPDSLVKOTHPNPLESLDLCAGAPLNOS	225			
QY	181	EVLAVGGSMIIIGIDHSLYTGSILWTPTRREMYEYVIVRAEINQDLKMDCKEKNYOK	240			
Db	226	EVLAVGGSMIIIGIDHSLYTGSILWTPTRREMYEYVIVRAEINQDLKMDCKEKNYOK	285			
QY	241	STVDSGTTMLRLPKRYEAAVAKSIIKAASSTEFKPDGFWLGEOLVCMQAGTTPWNIFFVIS	300			
Db	286	STVDSGTTMLRLPKRYEAAVAKSIIKAASSTEFKPDGFWLGEOLVCMQAGTTPWNIFFVIS	345			
QY	301	LYIMGVTVNOSRITLLPOOYLRPVEDVANTSDDDCKFPAISOSTGTWGAIVMEGFYV	360			
Db	346	LYIMGVTVNOSRITLLPOOYLRPVEDVANTSDDDCKFPAISOSTGTWGAIVMEGFYV	405			
QY	361	PPRRARRIGFAVSACHVHDEFTAAVEGFEYLLDMEDCGYNIPOTDSTLMTIAYVMAAI	420			
Db	406	PPRRARRIGFAVSACHVHDEFTAAVEGFEYLLDMEDCGYNIPOTDSTLMTIAYVMAAI	465			
QY	421	CALFMLPLCLAMCQWRCLRCRLROHDDFDADISLLK	456			
Db	466	CALFMLPLCLAMCQWRCLRCRLROHDDFDADISLLK	501			

```

1      RESULT 2
2      US-09-009-191-4
3      ; Sequence 4, Application US/09009191
4      ; Patent No. 6319689
5      ; GENERAL INFORMATION:
6      ; APPLICANT: POWELL, DAVID
7      ; APPLICANT: CHAPMAN, CONRAD
8      ; APPLICANT: MURPHY, RAY
9      ; APPLICANT: SMITH, TRUDI
10     ; TITLE OF INVENTION: ASP2
11     ; NUMBER OF SEQUENCES: 6
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: RATNER & PRESTIA
14     ; STREET: P.O. BOX 980
15     ; CITY: VALLEY FORGE
16     ; STATE: PA
17     ; COUNTRY: USA
18     ; ZIP: 19482
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Diskette
21     ; COMPUTER: IBM Compatible
22     ; OPERATING SYSTEM: DOS
23     ; SOFTWARE: FASTSO for Windows Version 2.0
24     ; CURRENT APPLICATION DATA:
25     ; FILING DATE: US/09/009,191
26     ; FILING DATE: 20-JAN-1998
27     ; CLASSIFICATION:
28     ; PRIORITY APPLICATION DATA:
29     ; APPLICATION NUMBER: UK 9701684.4
30     ; FILING DATE: 28-JAN-1997
31     ; ATTORNEY/AGENT INFORMATION:
32     ; NAME: PRESTIA, PAUL F
33     ; REGISTRATION NUMBER: 23,031
34     ; REFERENCE/DOCKET NUMBER: GH-70368
35     ; TELECOMMUNICATION INFORMATION:
36     ; TELEPHONE: 610-407-0700
37     ; TELEFAX: 610-407-0701
38     ; TELEX: 846169

```

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-009-191-4

```

Query Match	95.9%	Score 2320;	DB 4;	Length 774;
Best Local Similarity	98.9%;	Pred. No. 7.1e-241;		
Matches 439;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;

QY	13	GSFVEMVNLNRKSSQGGYYVEVMTGSSPQOTNLILVDTGSSNFVGAAPHPLHRYOROL	7
Db	1	GSFVEMVNLNRKSSQGGYYVEVMTGSSPQOTNLILVDTGSSNFVGAAPHPLHRYOROL	60
QY	73	SSTYDRLKGYVPTPGKMBGELGTDLVSTPHGNVTVRANIAITESDKFEFLNGSWE	13
Db	61	SSTYDRLKGYVPTPGKMBGELGTDLVSTPHGNVTVRANIAITESDKFEFLNGSWE	120
QY	133	GILGLAAVEIARPDLSLEPFDSLVKQTHVNLFSLQJCGAGFLPNQSEVLASVGGSMII	19
Db	121	GILGLAAVEIARPDLSLEPFDSLVKQTHVNLFSLQJCGAGFLPNQSEVLASVGGSMII	180
QY	193	GGIDHSLYTGLSWTPYIRREWEYEVIIYRVEINGODLMDCKEYWDKSIYDSGTTNLR	25
Db	181	GGIDHSLYTGLSWTPYIRREWEYEVIIYRVEINGODLMDCKEYWDKSIYDSGTTNLR	24
QY	253	PKRYVEAAVKSIIKASSTREKPPDGGWLEBOLVCMQAGTTPNNIPFVLSLYMGEYTNOSF	31
Db	241	PKRYVEAAVKSIIKASSTREKPPDGGWLEBOLVCMQAGTTPNNIPFVLSLYMGEYTNOSF	30
QY	313	RTTILPQOYLREVEVATSQDDCYFAAISQSSGTGYMAVIMEGYYVVEDRARRKIGFV	37
Db	301	RTTILPQOYLREVEVATSQDDCYFAAISQSSGTGYMAVIMEGYYVVEDRARRKIGFV	36
QY	373	SACHVDEERTAAVEGPFVTLDMEDCGINIPQTDSTLMTIATVMAAICALPMLPLCLMV	43
Db	361	SACHVDEERTAAVEGPFVTLDMEDCGINIPQTDSTLMTIATVMAAICALPMLPLCLMV	420
QY	433	COMRCLRCIRQOHDDEADISILK	456
Db	421	COMRCLRCIRQOHDDEADISILK	444

```

3
US-08-999-723-2
; Sequence 2, Application US/0899723A
; Patent No. 6025180
; GENERAL INFORMATION:
; APPLICANT: POWELL, David J.
; APPLICANT: Soutlan, Christopher
; APPLICANT: Chapman, Conrad G.
; APPLICANT: Evans, Joanne R.
; TITLE OF INVENTION: ASPI
; FILE REFERENCE: GH70262
; CURRENT APPLICATION NUMBER: US/08/999,723A
; CURRENT FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: prt
; ORGANISM: Homo sapiens
US-08-999-723-2

```

Query Match	48.4%;	Score 1171.5;	DB 3;	Length 518;
Best Local Similarity	49.6%;	Pred. No. 2.2e-117;		
Matches 222;	Conservative 80;	Mismatches 139;	Indels 7;	Gaps 4;

[illegible]

```

RESULT      4
US-09-434-427-2
: Sequence 2, Application US/09434427
: Patent No. 6162630
: GENERAL INFORMATION:
:   APPLICANT: POWELL, DAVID J.
:   APPLICANT: SOUTHAN, CHRISTOPHER
:   APPLICANT: CHAPMAN, CONRAD G.
:   APPLICANT: EVANS, JOANNE R.
:   TITLE OF INVENTION: ASPI
:   FILE REFERENCE: GH-70262-D1
:   CURRENT APPLICATION NUMBER: US/09/434,427
:   CURRENT FILING DATE: 1999-11-04
:   EARLIER APPLICATION NUMBER: US 08/999,723
:   EARLIER FILING DATE: 1997-10-06
:   EARLIER APPLICATION NUMBER: UK 9626022.9
:   EARLIER FILING DATE: 1996-12-14
:   NUMBER OF SEQ ID NOS: 2
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 2
:   LENGTH: 518
:   TYPE: PRT
:   ORGANISM: HOMO SAPIENS
:   US-09-434-427-2

```

[illegible]

Db 191 IKMNGILGAAVATLAKPSSSLETFEFDLSLYOTQANIPNVFSMOMCCAGLPAVGS---GTNGG 247

Qy 189 SMTIGGIDHSLYTGSLSMTPIPIREMYEVIIIVAEINGODLKMDKREYNDRKDSIVSGTT 248

Db 248 SLVIGGIEEPIISLYKGIOWTPIKEEMYYOIEILKELIGGOSLINDCREXNMADKAIIVSGTT 3077

Qy 249 NLRLEKRYFEAAVKSIRKAASSTEKFPDGFWMJGEOJYOMQAGTPPMNIFEPISILYNGEVT 308

Db 308 LLRLPKVFDAVVEAVARASLLPEFSDGFWNGSLDACWTSEIHPWISFPFISITYLRENS 367

Qy 309 NOSFEITILLPOOLRYVEDVATSDDDCKFAISQSSGTYGMVAMEGFEVYEDRARRI 368

Db 368 SSSFEITILLPOLYIQPMAGCLNT-ECYRFGISPTNALYIGATVMGFEVIEDRARRI 426

Qy 369 GFVAASCHVHDEFTAAVEGFEVTLMDKEDCGYNTPQTDSELTAVIYMAVAI-ALEMLP 427

Db 427 GFPAASCAEIGAIVASEISGPFSEEDVASNVCNPQASLSEPIIMIVSTALMSVCCAILLV 486

Qy 428 LCIAMVQMRCLRLROOHDPADIDISLL 455

Db 487 IYLLLLPFCQR--RPDRPEVYVNDSSSL 512

```

? RESULT 5
? US-09-717-432-2
? Sequence 2, Application US/09717432
? Patent No. 6291223
? GENERAL INFORMATION:
? APPLICANT: ZHU, YUAN
? APPLICANT: LI, XIAOTONG
? APPLICANT: CHRISTIE, GARY
? APPLICANT: POWELL, DAVID J.
? TITLE OF INVENTION: Mouse Aspartic Secretase-1 (masp1)
? FILE REFERENCE: GP-70663
? CURRENT APPLICATION NUMBER: US/09/717,432
? CURRENT FILING DATE: 2000-11-21
? PRIOR APPLICATION NUMBER: 60/166,974
? PRIOR FILING DATE: 1999-11-23
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 2
? LENGTH: 514
? TYPE: PRT
? ORGANISM: MUS MUSCULUS
? US-09-717-432-2

```

	Query Match	47.4%;	Score 1146.5;	DB 4;	Length 514;
	Best Local Similarity	50.7%;	Pred. No. 1.le-114;		
	Matches 216;	Conservative 75;	Mismatches 130;	Indels 5;	Gaps 3;
QY	8 EPRGRGSEVEMVDNLROCKSGCGYVEMTVSSPPTLNLIVDTGSNNFPAVGAPPHLR	66			
Dd	65 EPVATANAFMLAWMDIQDSDRGRYLEMLIGRPXQGIIVDTGSSNFAVAGAPHISYIDT	124			
QY	67 YYRQLSTTYADLKRGVYPVTQCKWBEELGTDLVISPHGPNTVRANIAATESDKFFI	126			
Dd	125 YFDSESSSTYSKGFDPVAVKYTGOSWTFEVEDLVTTIKFGNSSFLVINIAITFESENFPL	184			
QY	127 NGSNMGEIGILGLAYAEIARPDDSLIEPFDFSLVKOTHPVLPSLTOLCGAEPFNQSEVLASV	186			
Dd	185 PGIMWNGILGLAYALAKFPSSSLETFFPDSLVAAQAKIPIFSKMOCGAALPVAGS---GTN	241			
QY	187 GGSMTIGIGIDHSLYTGSLSWTPPIRREWEYEVIIVRVETINGODLKMDCKEYNVDKSIYDSC	246			
Dd	242 GGSLVLTGGIEPSLTKGDIMWTPPIKEEWYEQEIIKLFTIGCONLINDCREYNADRAIVDSC	301			
QY	247 TTNRLRKVEEAVALVKSIAKASTTEKPPDGWGMLEOOLVCMAQTGMPIIRFVISTLYLMGE	306			
Dd	302 TTLRLRLQKVDAVEAVARSLLPEPSSDGGWTGAQLCAWMNSSELPMAVFKRISITYLADE	361			
QY	307 VTNOSEFRITILLPOOYLRRVEDVAATSDDCYKFAISQSSTGTVMAGVINEGCVYVVEDFKAR	366			

```

Db 362 NASSFRITLLPOLXIQPMKAGCNT-ECRFGISSTNLVLGATVMEGFYVFPDQR 420
OY 367 RIGRAVSACHVHDEFRTAAYVGPVPLTDDMEDCGNINQGTDESTLMTATAYMAALCALEML 426
Db 421 RVRGASVSCAEIEGTIVSEISGPFSTEDIASNCYPAQALNEPILMTIVSYALMSVCGAILL 480
OY 427 PLCLTAV 432
Db 481 VLITLL 486

```

```

? RESULT 6
? US-09-912-484-2
? Sequence 2, Application US/09912484
? Patent No. 6358725
? GENERAL INFORMATION:
? APPLICANT: Christie, Gary
? APPLICANT: Li, Xiaotong
? APPLICANT: Powell, David J.
? APPLICANT: Zhu, Yuan
? TITLE OF INVENTION: Mouse Aspartic Secretase-1 (masp1)
? FILE REFERENCE: GP-70663-D1
? CURRENT APPLICATION NUMBER: US/09/912,484
? CURRENT FILING DATE: 2001-07-25
? PRIOR APPLICATION NUMBER: 60/166,974
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: 09/717,432
? PRIOR FILING DATE: 2000-11-21
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 2
? LENGTH: 514
? TYPE: PRP
? ORGANISM: MUS MUSCULUS
? US-09-912-484-2

```

Query Match	47.48;	Score 1146.5;	DB 4;	Length 514;
Best Local Similarity	50.78;	Pred. No. 1.1e-14;		
Matches 216;	Conservative	130;	Indels	5;
			Gaps	3

[illegible]

```

1 RESULT 7
2 US-08-208-007A-13
3 ; Sequence 13, Application us/08208007A
4 ; Patent No. 5501969
5 ;
6 ; GENERAL INFORMATION:
7 ; APPLICANT: HASTINGS, ET AL.
8 ; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
9 ; NUMBER OF SEQUENCES: 14
10 ; CORRESPONDENCE ADDRESSES:
11 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
12 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN
13 ; STREET: 6 BECKER FARM ROAD
14 ; CITY: ROSELAND
15 ; STATE: NEW JERSEY
16 ; COUNTRY: USA
17 ;
18 ; ZIP: 07068
19 ;
20 ; COMPUTER READABLE FORM:
21 ; MEDIUM TYPE: 3.5 INCH DISKETTE
22 ; COMPUTER: IBM PS/2
23 ; OPERATING SYSTEM: MS-DOS
24 ; SOFTWARE: WORD PERFECT 5.1
25 ;
26 ; CURRENT APPLICATION DATA:
27 ; APPLICATION NUMBER: US/08/208,007A
28 ; FILING DATE: March 8, 1994
29 ; CLASSIFICATION: 435
30 ;
31 ; PRIOR APPLICATION DATA:
32 ; APPLICATION NUMBER: No. 5501969e
33 ; FILING DATE: No. 5501969e
34 ; ATTORNEY/AGENT INFORMATION:
35 ; NAME: FERRARO, GREGORY D.
36 ;
37 ; REGISTRATION NUMBER: 36,134
38 ; REFERENCE/DOCKET NUMBER: 352800-95
39 ;
40 ; TELECOMMUNICATION INFORMATION:
41 ; TELEPHONE: 201-994-1700
42 ; TELEFAX: 201-994-1744
43 ;
44 ; INFORMATION FOR SEQ ID NO: 13:
45 ; SEQUENCE CHARACTERISTICS:
46 ; LENGTH: 396 AMINO ACIDS
47 ; TYPE: AMINO ACID
48 ; STRANDEDNESS:
49 ; TOPOLOGY: LINEAR
50 ;
51 ; MOLECULE TYPE: PROTEIN
52 ;
53 US-08-208-007A-13

```

Query Match	12.4%	Score 299.5	DB 1	Length 396
Best Local Similarity	25.9%	Pred. No. 1.3e-23		
Matches 100; Conservative	65;	Mismatches 148;	Indels 73;	Gaps 16

```

OY 3 DEBPEEGRGRSEVEMVNDLRRKSGQGYVVEWPGSPOTLNLIVDGSNIFAGA---- 58
Db 63 DOSAKP-----LINTLD-----MEFPIITISIPPOONTVFIEDIGSSNLWPVSYCT 110
OY 59 APPELHRYYQOLSTSYDLRKGYVYPTOGKWEGLCTDLYSLPIGPNVYRANIAAI 118
Db 111 SPACKTHSRFOSOSSTSYQOPQSFSTIYGTGSLGILGAQVSV-BELTVGGQSESV 169
OY 119 TEEDKRFINGSWMEGLTGLAYAEIARPDSDLEPPEDSLVYKOTHVNFSLDLCGAGPLN 178
Db 170 TEGGQGFVVD-AEEDGILGIGYPSLA--VGVYTPVVDNMMAC-----NLVDLEPMFSVMSSN 222
OY 179 OSEVYLAVSGSMITGDISLSTGSLWTPPIRREKXYEVIIIVRVEINGQDLKMKCKEYNY 238
Db 223 PE--GAGSELLEFGGYDHSHPGSSLNWPVKQIMQJALDNLNIOVGS--IVMFCSE--G 275
OY 239 DKSIVSDGNTLRLPKRVEEAHVSIKASSTEKPEPDGFWLGEOLVQMOAGTPWNIIPV 298
Db 276 CQALIVDTGSLTILGSPDKIKQLONALIGAAP-----VDEEYAVE-----CANLNVMPD 322
OY 299 ISLYLGEVYNOSFRITLIPQOYLARVEVEVATSDQCKRFALSSOSTG----- 346
Db 323 VFFPINS-----VPLYLSTJAL--TLIDPVDGMOFC-----SSSGFOGLDHPAGP 366

```

QY 347 -TVMGAVIMEGFYVFDARRKRTGFA 371  
Db 367 LMIIGDVFIROFYSVFDGNNRVGLA 392

RESULT 8  
US-09-032-523-9

; Sequence 9, Application US/09032523  
; Patent No. 6232454  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guebler, Karl  
; APPLICANT: Baugh, Mariah  
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/032,523  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0479 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 181994  
; US-09-032-523-9

Query Match 12.4%; Score 299.5; DB 4; Length 396;  
Best Local Similarity 25.9%; Pred. No. 1.3e-23;  
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

QY 3 DEEPEPGRGSEVEMVNLNRKSGQGYVEMTVGSPQTLNIVDTGSSNFAVGA----- 58  
Db 63 DOSAKEP-----LINVLD-----MEYFGTISTISPPONTVFIFDTGSSNLMWPSYYCT 110  
QY 59 AHPRLHYYQRLSSTYRDLRKGYVYPTGQKMEGELGTLVSIPIGPNVTVRANITAI 118  
Db 111 SPACTHRRFPQSSSTYSQPGQSFSTIOYGTGSGITIGADQVSV-BGLTVVGGQFGESV 169  
QY 119 TESDFKFFSNMGEIILGLAVETARPDSDLPEPFDLSLVKQTHVNFSLDLCAGFPLN 178  
Db 170 TEPGQTFVD-AEPDGLGLGTPSLA--VGVVTVFVDNMAO-----NLVDLPMSFSVYSSN 222

QY 179 QSEVLASVGSMIIGSDHSLYTSIMYTPPIREMYEYIIVREINODLMDCKEYNY 238  
Db 223 PE---GGASSELPFGYDHSHPSSGLNWPYTKQAYWJALDNIQVG--TVMFCSE--G 275  
QY 239 DKSIVDSGTNLRPKVFEAAVKSIAKASTKEKPPDGFMLGEOQLVCMQAGTTPNIPV 298  
Db 276 COAIVDTGTSILTPSDKIKQLQNAIGAP-----VDEYANE-----CANLVMPD 322  
QY 299 ISLYLMGEVTNOSFRITLIPQOYLRPVEDVATSDQCYKFAISQSTG----- 346  
Db 323 VFTTNG-----VPTLSPTAY--TLDFVDGMQFC-----SSGFGGLDIHPAPG 366  
QY 347 -TVMGAVIMEGFYVFDARRKRTGFA 371  
Db 367 LMIIGDVFIROFYSVFDGNNRVGLA 392

RESULT 9  
US-08-208-007A-12

; Sequence 12, Application US/08208007A  
; Patent No. 5501969  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESS: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/208,007A  
; FILING DATE: March 8, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: No. 5501969e  
; FILING DATE: No. 5501969e  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 412 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; US-08-208-007A-12

Query Match 12.2%; Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. No. 3.8e-23;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;

QY 30 YVEMTVGSPQTLNIVDTGSSNFAVGAAPHEL-----HRYQROLSTYRDLRKGV 83  
Db 79 YVEIGIGTPPOCFVTVPDTGSSNLMWPSIHCKLIDACMIHKKVYKSDKSTYYVNGTSF 138  
QY 84 YVPTGQKMEGELGTLVSIPIGPNVTVRANITAIRESKFFINGSNMGEI 134  
Db 139 DIHYGSGSLSGTISQDTVAVPCOSASSASALGVKVERQVFEATRKQPGITFIAPKFDGI 198

OY	135	LGLVAEIAIARDPDSLEFPEDSLVKQTHV-PNLSLOJCGAGFPLNOSEVLASVGSSMIIG	193
		:     :     :     :     :	
Dd	199	LGMAYPRIS--VNVTLEVPDMLMOOKLVDONISEFYI-----SKDPDAOPGEGELMIG	248
OY	194	GIDASLTGSLSWYPPIREWEVEYIIIVRVEI-NSODLKMDCKEYNKSDSVDSGTTLRL	252
		: : : : :       :       :       :       :	
Dd	249	GTDSEKYKXGSLSYLNVRKKAWOYHLDPOVEVASGLTFL---CKE--GCBAIVDTGTSIMVG	303
OY	253	PKKVPEAAVKISAASSTEFKPPDFWGEOLV-CMOAGTFPMWNIFPVISLYLMGEVYNQS	311
		:       :       :       :       :       :	
Dd	304	PVDEVELRKAIKAIGNAPLTO-----GEYNIPEKYST----LPRLITLKG-----KG	346
OY	312	FRIITLLPOOLLRPEDVATSODDCYKEAISO-----SSTGVMGAVIMEGYVFEDARK	366
		: : : :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	347	YKLS--PEDYTLLKVSQAOKTL-CLSGFMGDIPPPSGPLMIILGDVFGIRYTVFDKRON	402
OY	367	RIGFAVSA	374
		:       :	
Dd	403	KVGFEAEA	410



QY 261 VKSIIKASSTKFFPDGFWLGE-QLVCMQAGTTNNINPVLSTLMGEVTVNQSRITTLPO 319  
DQ 368 QQAIGATGNG-----XGEFDIDCNLSMPVIV-----ETCKMPLT--PS 408  
QY 320 QYLKRPVEVLAISODCY--KFAISQSTGTGKAGVIMESGFYVDFRRKRIGFA 371  
DQ 409 AY-----TSQAGCFTSGFQSEHNHQQWILLADVEITREYVSFPRANNLVGLA 455

```

1      RESULT 12
2      US-09-640-305-6
3      ; Sequence 6, Application US/09640305
4      ; Patent No. RE37447
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Fleer, Reinhard
7      ; Fournier, Alain
8      ; Yeh, Patrice
9      ;
10     TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
11     PREPARATION AND USE
12     ;
13     NUMBER OF SEQUENCES: 17
14     ;
15     CORRESPONDENCE ADDRESS:
16     ;
17     ADDRESSEE: Rhone-Poulenc Rorer Inc.
18     ;
19     STREET: 500 Arcola Rd. 3C43
20     ;
21     CITY: Collegeville
22     ;
23     STATE: PA
24     ;
25     COUNTRY: USA
26     ;
27     ZIP: 19002
28     ;
29     COMPUTER READABLE FORM:
30     ;
31     MEDIUM TYPE: floppy disk
32     ;
33     COMPUTER: IBM PC compatible
34     ;
35     OPERATING SYSTEM: PC-DOS/MS-DOS
36     ;
37     SOFTWARE: PatentIn Release #1.0, Version #1.25
38     ;
39     CURRENT APPLICATION DATA:
40     ;
41     APPLICATION NUMBER: US/09/640,305
42     ;
43     FILING DATE: 16-Aug-2000
44     ;
45     PRIOR APPLICATION DATA:
46     ;
47     APPLICATION NUMBER: US/08/360,673
48     ;
49     FILING DATE: 06-FEB-1995
50     ;
51     APPLICATION NUMBER: WO PCT/FR93/00623
52     ;
53     FILING DATE: 23-JUN-1993
54     ;
55     APPLICATION NUMBER: FR 92/07785
56     ;
57     FILING DATE: 25-JUN-1992
58     ;
59     ATTORNEY/AGENT INFORMATION:
60     ;
61     NAME: Smith, Julie K.
62     ;
63     REGISTRATION NUMBER: 38,619
64     ;
65     REFERENCE/DOCKET NUMBER: ST92040-US
66     ;
67     TELECOMMUNICATION INFORMATION:
68     ;
69     TELEPHONE: (610)454-3839
70     ;
71     TELEFAX: (610)454-3808
72     ;
73     INFORMATION FOR SEQ ID NO: 6:
74     ;
75     SEQUENCE CHARACTERISTICS:
76     ;
77     LENGTH: 409 amino acids
78     ;
79     TYPE: amino acid
80     ;
81     TOPOLOGY: linear
82     ;
83     MOLECULE TYPE: protein
84     ;
85     SEQUENCE DESCRIPTION: SEQ ID NO: 6:
86     ;
87     US-09-640-305-6

```

	Query Match	11.3%	Score 273.5;	DB 1;	Length 409.
	Best Local Similarity	25.3%;	Pred. No. 8.7e-21;		
Matches	92; Conservative	59;	Mismatches 144;	Indels	Gaps 11.
OY	30 YVETVGSPPTLLILVDTGSSNFAVCAP----	HPLHRYXOROLSTYRDLKGYV	85		
Db	: :-  ::    :	:::   :	:	::   :	
	96 YFTFILLPSQPFVVLDGTGSSNLMPVASAEGLSLACFLHTFYDHEASTTYANGSEPAI		155		
OY	86 PYTOKWEGELCTDYLSIPHGPNVTYVRANIAITSDKFIFNGSMWGIIIGLVAEIAEP		145		
Db	:  :   ::  :::				
	156 QYGSSLLEGIVSRDLLTI--GDLYLPDDDFEAATSEPGEAFAGFKFDGIIGLATAYDIS-V		212		
OY	146 DSIIEPFIIDSVLKQTTHVNVLISQLCGAGFPINOSEVLAVASGSMTIGDIHSITGSLM		205		
	:   : ::    :				

```

Db      213  NRIVPPYNAIKNKLLDDPVFA-----FLIGSD- KSEGGGASRGIGIDEEKVYGEIT 264
QY      206  YPIPIREMYEVIIVRVLTINQDILKMDCKEYNYDRSJIVDSGTNLRLPKRVFEAAVSIK 265
      265  WLPVRKAYWE-----VFEGIGIGEEFATLEHGAAIDTGTSLALPSSGLAELNNEIG 319
QY      266  AASSTKEKPPDGFMLGEOLVCMQATTPPINIFPVVSLYLMEVYNQSRRT----- 315
      320  AKKG-----WSQSVSDCESRDS---LPDLTTLFNFG---YNNFTIAYDYTLLEVSG 363
Db      320  AKKG-----WSQSVSDCESRDS---LPDLTTLFNFG---YNNFTIAYDYTLLEVSG 363
QY      316  -----ILPQQLRLRPEDVATVSQDDCYKFAISQSSGTGVMGAVIMEGFVYVEDBARRIGF 370
      364  SCISAFETPMDEPEVPGLA-----ITGDALRKRYSTLYDIGHNAVGL 405
QY      371  AVSA 374
      406  AKAA 409
Db      406  AKAA 409

```

```

1      RESULT 3
2      US-08-360-673-6
3      : Sequence 6, Application US/08360673
4      : Patent No. 5679344
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Fleer, Reinhard
8      : APPLICANT: Fournier, Alain
9      : APPLICANT: Yeh, Patricia
10     :
11     : TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEAST
12     : TITLE OF INVENTION: PREPARATION AND USE
13     :
14     : NUMBER OF SEQUENCES: 17
15     : CORRESPONDENCE ADDRESS:
16     : ADDRESSER: Rhone-Poulenc Rorer Inc.
17     : STREET: 500 Arcola Rd. 3C43
18     : CITY: Collegeville
19     :
20     : STATE: PA
21     :
22     : COUNTRY: USA
23     :
24     : ZIP: 15002
25     :
26     : COMPUTER READABLE FORM:
27     :
28     : MEDIUM TYPE: Floppy disk
29     : COMPUTER: IBM PC compatible
30     : OPERATING SYSTEM: PC-DOS/MS-DOS
31     : SOFTWARE: PatentIn Release #1.0, Version #1.25
32     :
33     : CURRENT APPLICATION DATA:
34     : APPLICATION NUMBER: US/08/360,673
35     :
36     : FILING DATE:
37     :
38     : CLASSIFICATION: 435
39     :
40     : PRIOR APPLICATION NUMBER:
41     : APPLICATION NUMBER: WO PCT/FR93/00623
42     :
43     : FILING DATE: 23-JUN-1993
44     :
45     : PRIOR APPLICATION DATA:
46     : APPLICATION NUMBER: FR 92/07785
47     : FILING DATE: 25-JUN-1992
48     :
49     : ATTORNEY/AGENT INFORMATION:
50     : NAME: Smith, Julie K.
51     :
52     : REGISTRATION NUMBER: 38,619
53     :
54     : REFERENCE/DOCKET NUMBER: ST92040-US
55     :
56     : TELECOMMUNICATION INFORMATION:
57     : TELEPHONE: (610)454-3839
58     : TELEFAX: (610)454-3808
59     :
60     : INFORMATION FOR SEQ. ID NO: 6:
61     :
62     : SEQUENCE CHARACTERISTICS:
63     :
64     : LENGTH: 409 amino acids
65     :
66     : TYPE: amino acid
67     :
68     : TOPOLOGY: linear
69     :
70     : MOLECULE TYPE: protein
71     :
72     : US-08-360-673-6

```

Query Match	11.3%	Score 273.5;	DB 1;	Length 409;
Best Local Similarity	25.3%	Pred. No. 8.7e-21;		
Matches	92;	Conservative 59;	Mismatches 144;	Indels 69;
			Gaps	11;

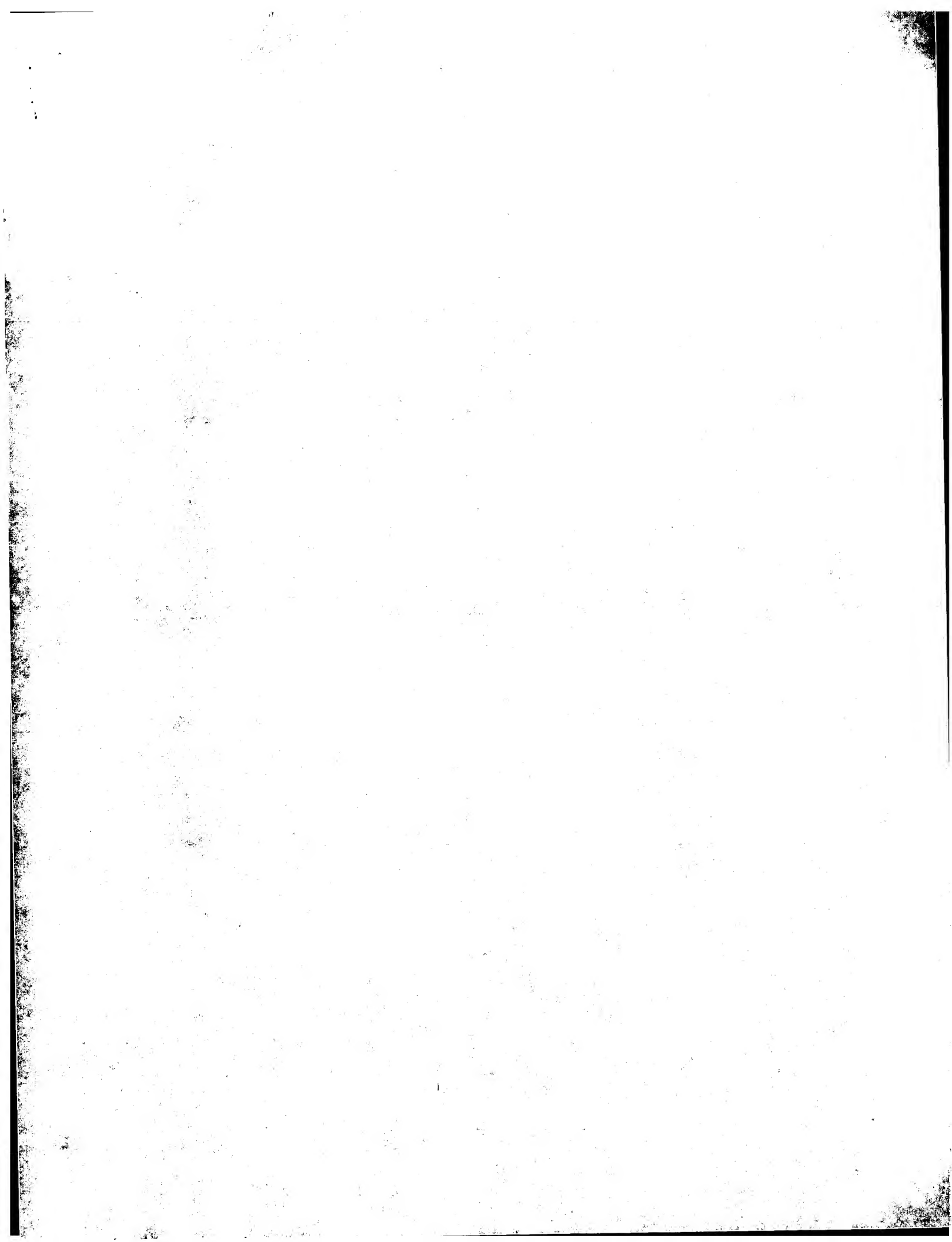


```

0Y 30 YVEVETVSSPQPTLIIIVDTSSSNVAVA-----APHPLHXYORQIOLSSTRYRDLRKGVV 85
Db 96 YTEVETSLGTPPQSFVILIDTSSNLMWPKSCGGSLACILHAKHYHDSSYTKKNGSSFEI 155
0Y 86 PYTQKMGCELTDLVSLPHGPNVTVRANIAITESDKFFNGSNMGGIIGLAVAEIARP 145
Db 136 RGGSSSMGGVYVSOVDLQI--GDLTTPKDFEAEATSEGLAFAPFGKFDGTIGLAV----- 207
0Y 146 DDLSEPFEDSLVKQTHVNLPL---SLOJCGA---GFLPNQSEVLASVGGSMIIIGIDHSL 199
Db 208 -----DSISVKKIYVPPIKALELDBDEPKRAFLIGTDKDESDGGATGEGVYKSK 259
0Y 200 YTGSLMTPIRREMYEVYIYRVINODLKMCKEYNYDKSIVDSGTTMLRLPKKYFEA 259
Db 260 YEGGITHLMPARKKAYWE-----VSFDEVGAGSEVAELQKGAIDTGTSLIALPSGIAEI 314
0Y 260 AVKSIRKAASSTEKEPDDGFMLGEOLVCMQAGTTPWNIPIFSLVLMGEVNTQNSRITTLRP 319
Db 315 LMAEIGATKG-----MSGQYAVDDCDTRDS-----LPDLTLFPAG-----YNFTITTP 356
0Y 320 QYLEPVDVATSOODCYKFAISQSSSTGTVCANVIMEGFYVPEPRARRKRGFAVS 373
Db 357 DYTELVSGGSCISATPMDPF--EPRLGPIALIGDSFLRKRYISVYDLGDAVGLAAS 409

```

```
Search completed: August 7, 2002, 09:14:48
Job time: 73 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:13:35 ; Search time 56.55 Seconds  
(without alignments)  
895.661 Million cell updates/sec

Title: US-09-724-571-43  
Perfect score: 2419  
Sequence: 1 ETDEEPEEPRGRSGFEWMD.....CLRLRQOHDFADISLTK 456

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

\_A\_Geneseq\_032802:\*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2419	100.0	456	21	AAB07897 Active enzyme port
2	2419	100.0	488	22	AAB66572 Human memapsin 2.
3	2419	100.0	488	22	AAB61334 Memapsin 2 protein
4	2419	100.0	501	21	AA94767 Human beta-secretase
5	2419	100.0	501	21	AAB07896 Human beta-secretase
6	2419	100.0	503	22	AAB6573 Human pro-memapsin
7	2419	100.0	503	22	AAB61335 t7 promoter and ve
8	2414	99.8	501	21	AA588425 Human aspartyl pro
9	2414	99.8	501	22	AAE10629 Human aspartyl pro
10	2414	99.8	501	22	AAE06859 Human aspartyl pro
11	2414	99.8	501	22	AAU06603 Human Aspartyl pro

12	2414	99.8	501	22	AAU07202 Human aspartyl pro
13	2414	99.8	501	22	AAE02581 Human aspartyl pro
14	2413	99.8	501	19	AAW59807 Amino acid sequenc
15	2392	98.9	501	21	AA94769 Rat beta-secretase
16	2390	98.8	501	21	AA94768 Murine beta-secret
17	2390	98.8	501	21	AAV88427 Murine aspartyl pr
18	2390	98.8	501	22	AAE10631 Murine aspartyl pr
19	2390	98.8	501	22	AAE06861 Murine aspartyl pr
20	2390	98.8	501	22	AAU06605 Mouse Aspartyl pro
21	2390	98.8	501	22	AAU07204 Mouse aspartyl pro
22	2390	98.8	501	22	AAE04948 Mouse aspartyl pro
23	2390	98.8	501	22	AAE02583 Mouse aspartyl pr
24	2351.5	97.2	969	22	ABG09611 Novel human diago
25	2320	95.9	790	19	AAW59808 Partial amino acid
26	2264.5	93.6	476	21	AAV88426 Human aspartyl pro
27	2264.5	93.6	476	22	AAE10630 Human aspartyl pro
28	2264.5	93.6	476	22	AAE06860 Human aspartyl pro
29	2264.5	93.6	476	22	AAU06604 Human Aspartyl pr
30	2264.5	93.6	476	22	AAU07203 Human aspartyl pro
31	2264.5	93.6	476	22	AAE02582 Human aspartyl pr
32	2238.5	92.5	476	22	AAE06909 Murine aspartyl pr
33	2238.5	92.5	476	22	AAE02619 Murine aspartyl pr
34	2212	91.4	427	22	AAV93866 Human polypeptide,
35	2160	89.3	453	21	AAV88438 Modified human asp
36	2160	89.3	453	22	AAE10642 Human-Asp 2(a) del
37	2160	89.3	453	22	AAE06872 Human-Asp 2(a) del
38	2160	89.3	453	22	AAE06616 Human-pro-Asp 2(a)
39	2160	89.3	453	22	AAU07215 Human aspartyl pro
40	2160	89.3	453	22	AAE02594 Human-Asp 2(a) del
41	2160	89.3	459	21	AAV88439 Modified human asp
42	2160	89.3	459	22	AAE10643 Human-Asp 2(a) pro
43	2160	89.3	459	22	AAE06873 Human-Asp2(a) del
44	2160	89.3	459	22	AAU06617 Human-pro-Asp 2(a)
45	2160	89.3	459	22	AAU07216 Human aspartyl pro

#### ALIGNMENTS

RESULT 1

AAU07897 standard; Protein; 456 AA.

AC AAB07897;

14-NOV-2000 (first entry)

DE Active enzyme portion of human beta-secretase enzyme.

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;

KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

KW inhibitor; ss.

OS Homo sapiens.

XX WO200047618-A2.

PN

XX

PD 17-AUG-2000.

XX

PF 10-FEB-2000; 2000WO-US03819.

XX

PR 10-FEB-1999; 99US-0119571.

XX

PR 15-JUN-1999; 99US-0139172.

XX

PA (ELAN) ELAN PHARM INC.

XX

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

XX Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

DR WPI: 2000-533011/48.

XX

PT Purified beta-secretase protein used in assays to discover inhibitors

PT which can be used for the treatment of amyloidogenic diseases e.g.

PT Alzheimer's disease -  
 XX  
 PS Claim 24; Fig 2b; 121pp; English.  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents the active enzyme portion of human  
 CC beta-secretase enzyme.  
 CC  
 XX Sequence 456 AA;  
 SQ  
 Query Match 100.0%; Score 2419; DB 21; Length 456;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-242; Indels 0; Gaps 0;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ETDEPEEPGRGSGFVEMVDNLKSGSGGYVEMTVGSPPTLNILVDTGSSNFAVGAP 60  
 DB 1 etdeepeepgrgsgfvevmdnlrgsgggyvemtvgspptlnilvdtgsnfavgap 60  
 QY 61 HPFLHRYQROLSTYRDLRKGVYPTQKGEGELGTDLVSIHPGPNTVANAIAITE 120  
 DB 61 hpflhryqrolstyrdlrkgyvpytqkgegelgtdlvsihpgpntvranaiate 120  
 QY 121 SDRFFINGSNMGEIIGLVAEIAEPDSDLPEPFDLSLVKOTHPNLFSLQCGAGPELNS 180  
 DB 121 sdrffingsnmgeiiglvaeiaepdSDLPEPFDLSLVKOTHPNLFSLQCGAGPELNS 180  
 QY 121 sdrffingsnmgeiiglvaeiaepdSDLPEPFDLSLVKOTHPNLFSLQCGAGPELNS 180  
 DB 121 sdrffingsnmgeiiglvaeiaepdSDLPEPFDLSLVKOTHPNLFSLQCGAGPELNS 180  
 QY 181 EYLASVGSMTIIGIDHSLYTGSMTPTIRREMYEVIIIVRVEINGODIKMCKEYNDK 240  
 DB 181 eylvsgsmtiigidhsllytgsmtptirremyeviiivrveingodikmckeyndk 240  
 QY 241 SYVDGSTNLRLPKKVFEEAAVKSIAASTKEKPPDGFMLGEOLYCQAGTTPMNIPEVVS 300  
 DB 241 syvdgstnlrlpkkvfeaaavksiaastkeKPPDGFMLGEOLYCQAGTTPMNIPEVVS 300  
 QY 301 LYIMGEVNTQSFRTITLPQOYLRPVEDVATSDDCYKFAISOSTGTWGAIVMEGFYVV 360  
 DB 301 lyimgevtngsfrtitlpqoylrpvedvatSDDCYKFAISOSTGTWGAIVMEGFYVV 360  
 QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESTLMTIAYMAAI 420  
 DB 361 fdrarrkigfavsachvhdefrtaavegpfvtldmedcgynipqdestlmtiaymaai 420  
 QY 421 CALFMLPICLMWCQWRCLRCLRQOHDPADDISILK 456  
 DB 421 calfmplclmwcwrcrlrcrlrqohdtdaddisilk 456

RESULT 2

AAB66572

ID AAB66572 standard: Protein; 488 AA.

AAB66572;

12-APR-2001 (first entry)

Human memapsin 2.

Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;

APP; memapsin 2 inhibitor; Alzheimer's disease.

Homo sapiens.

WO200100665-A2.

XX

PD 04-JAN-2001.  
 XX  
 XX 27-JUN-2000; 2000WO-US17742.  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA (UNIT ) UNITV ILLINOIS FOUND.  
 PI Tang JUN, Hong L, Ghosh AK;  
 DR WPI: 2001-137933/14.  
 DR N-PSDB; AAF31848.  
 PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage  
 XX  
 PS Example 1; Page 72-74; 86pp; English.  
 XX  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 XX Sequence 488 AA;  
 SQ  
 Query Match 100.0%; Score 2419; DB 22; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-242;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ETDEPEEPGRGSGFVEMVDNLKSGSGGYVEMTVGSPPTLNILVDTGSSNFAVGAP 60  
 DB 33 etdeepeepgrgsgfvevmdnlrgsgggyvemtvgspptlnilvdtgsnfavgap 92  
 QY 61 HPFLHRYQROLSTYRDLRKGVYPTQKGEGELGTDLVSIHPGPNTVANAIAITE 120  
 DB 93 hpflhryqrolstyrdlrkgyvpytqkgegelgtdlvsihpgpntvranaiate 152  
 QY 121 SDRFFINGSNMGEIIGLVAEIAEPDSDLPEPFDLSLVKOTHPNLFSLQCGAGPELNS 180  
 DB 153 sdrffingsnmgeiiglvaeiaepdSDLPEPFDLSLVKOTHPNLFSLQCGAGPELNS 212  
 QY 181 EYLASVGSMTIIGIDHSLYTGSMTPTIRREMYEVIIIVRVEINGODIKMCKEYNDK 240  
 DB 213 eylvsgsmtiigidhsllytgsmtptirremyeviiivrveingodikmckeyndk 272  
 QY 241 SYVDGSTNLRLPKKVFEEAAVKSIAASTKEKPPDGFMLGEOLYCQAGTTPMNIPEVVS 300  
 DB 273 syvdgstnlrlpkkvfeaaavksiaastkeKPPDGFMLGEOLYCQAGTTPMNIPEVVS 332  
 QY 301 LYIMGEVNTQSFRTITLPQOYLRPVEDVATSDDCYKFAISOSTGTWGAIVMEGFYVV 360  
 DB 333 lyimgevtngsfrtitlpqoylrpvedvatSDDCYKFAISOSTGTWGAIVMEGFYVV 392  
 QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESTLMTIAYMAAI 420  
 DB 393 fdrarrkigfavsachvhdefrtaavegpfvtldmedcgynipqdestlmtiaymaai 452  
 QY 421 CALFMLPICLMWCQWRCLRCLRQOHDPADDISILK 456  
 DB 453 calfmplclmwcwrcrlrcrlrqohdtdaddisilk 488

```
RESULT 3
AAB61334
ID AAB61334 standard; protein: 488 AA.
XX
AC AAB61334;
XX
DT 02-APR-2001 (first entry)
XX
DE Memapsin 2 protein.
XX
KM Memapsin 2; catalyst; Alzheimer's.
XX
OS Homo sapiens.
XX
PN WO200100663-A2.
XX
PD 04-JAN-2001.
XX
PF 27-JUN-2000; 2000MO-US17661.
XX
PR 28-JUN-1999; 99US-0141363.
PR 30-NOV-1999; 99US-0168060.
PR 25-JAN-2000; 2000US-0177836.
PR 27-JAN-2000; 2000US-0176368.
PR 08-JUN-2000; 2000US-0210292.
XX
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX
PI Tang JUN, Lin X, Koelsch G;
XX
DR WPI: 2001-102885/11.
XX
PT Purified recombinant catalytically active memapsin 2, used to screen
PT inhibitors of it, which are used to treat and prevent Alzheimer's
XX disease -
XX
PS Claim 2; Page 73-75; 86pp; English.
XX
CC The present invention relates to a purified recombinant
CC catalytically active memapsin 2. The invention may be used for
CC isolating inhibitors which are used to treat or prevent
CC Alzheimer's disease. The invention may also be used to screen
CC for individuals more genetically prone to develop Alzheimer's
CC disease.
XX
SQ Sequence 488 AA:
```

```
Query Match 100.0%; Score 2419; DB 22; Length 488;
Best Local Similarity 100.0%; Pred. No. 4.3e-242;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ETDEEPEERGRGRSFVEMVDNLGRKSGGYVEMVGSPPOTLNTLVDTGSSNFAVGAAP 60
DB |||||||
OY 33 etdeepeegrgrgrsfvemdnlgrksggyvemtvgspptlntlvdtgssnfavgaap 92
DB |||||||
OY 61 HPELRRYQROQLSSTRYRDLRKGVYPTQGWEGELGTDLVSIHPGNVTVRANIAATTE 120
DB |||||||
OY 93 hpeLrryqrqlsstyrldlrkyvypyrqgwegelgtdlvsihpgnvtvraniaalte 152
DB |||||||
OY 121 SDFEFINSNNEGIGLGYAEIARPDSDLPEFPISIVKQTIVPNLFSQLCGAGFPLNQS 180
DB |||||||
OY 153 sdfefinsnegiiglgyaeiarpdSDLPEFPISIVKQTIVPNLFSQLCGAGFPLNQS 212
DB |||||||
OY 181 EVLASVGSGMTIGGIDHSLYTGSLMYTPPIREWEYEVITIVAEINGODLKMDCKEYNVDK 240
DB |||||||
OY 213 evlasvgsgmtiggidhsllytgslmytppireweyevitivaeingodlkmdeckeynydk 272
DB |||||||
OY 241 SIYVSGTNNLPLPKKVFPAAKSKTKAASSTKFPDGFNLGQQLVCWQAGTTPWNIFFVIS 300
DB |||||||
OY 273 slvsgtntlplpkkvfpaaaksktkaaSSTKfPDgfnlGsqQLvcwqagTtpwnlffvis 332
```

```
OY 301 LYLMGEVTNOSRITITLPQOYLRPVEDVATSQDDCYKFAISOSSTGTGVKAVIMEGFYVY 360
DB |||||||
OY 333 lylmgevtnosrfititlpqoylrpvedvatSQDDCYKfAISOSSTGTGVKAVIMEGFYVY 392
DB |||||||
OY 361 FDRARRIGFAVSACHVHDEPRTAAVEGPFVTLDMEDCGYVNPQDESLMTIAVMAAI 420
DB |||||||
OY 393 fdrarrigfavsachvhdepRTAAVEGPFVTLdMEDcGYNpQDEsLMTIAVMAAI 452
DB |||||||
OY 421 CALFMLPLCLMWQMRCLRCLRQOHDFADDISLKL 456
DB |||||||
OY 453 calfmplclmwqmrclrclrqohdfaddislkl 488

RESULT 4
ID AAY94767 standard; Protein: 501 AA.
XX
AC AAY94767;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human beta-secretase amino acid sequence.
XX
KW Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; human;
KW Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Peptide 1..45
FH /label= putative signal peptide
FT Protein 46..501
FT /label= Beta-secretase
XX
EN WO200058479-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000MO-US07755.
XX
PR 26-MAR-1999; 99US-0277229.
XX
PA (AMGE-) AMGEN INC.
XX
PI Cytroon M, Vassar RJ, Bennett BD;
XX
DR WPI: 2000-594643/56.
DR N-PSDB: AAA28278.
XX
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful
PT for diagnosis and gene therapy of Alzheimer's disease -
XX
PS Claim 1; Fig 4; 145pp; English.
XX
CC This invention relates to 3 nucleotide sequences encoding beta-secretase
CC proteins. Beta-secretase is an enzyme involved in the production of one
CC of the components of amyloid plaques involved in Alzheimer's disease. The
CC invention includes an expression vector comprising the nucleotide
CC sequence, a host cell comprising the expression vector, and a process for
CC producing the protein through culturing the transformed cells. Also
CC included in the invention are a polypeptide derivative of the
CC beta-secretase protein, a fusion protein comprising beta-secretase fused
CC to a heterologous amino acid sequence, and a method for modulating the
CC levels of beta-secretase polypeptide in a mammal comprising administering
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
CC neurotrophic activity. The beta-secretase nucleotide sequence may be used to
CC map locations of the beta-secretase gene and related genes on chromosomes
CC and as hybridization probes in diagnostic assays to test for the presence
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
CC used as anti-sense inhibitors of beta-secretase expression, in gene
CC therapy of Alzheimer's disease, and for the identification of compounds
```

CC that modulate beta-secretase activity. Antibodies to the beta-secretase  
CC protein may be used for in vitro and in vivo diagnostic purposes to  
CC detect the presence of beta-secretase polypeptide in a body fluid or cell  
CC sample. The present sequence represents the human beta-secretase protein.  
CC  
XX  
SO Sequence 501 AA;

Query Match 100.0%; Score 2419; DB 21; Length 501;  
Best Local Similarity 100.0%; Pred. No. 4.5e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSPFVEMVDNLKRGSGGYVEMTVGSPPTLNIIVDTGSSNFVAGAAP 60  
DB 46 etdeepeepgrgsfvevmdnlrksggyvemtvgspptlnllydtgssnfavgaap 105  
QY 61 HPFLHRYQROLSTSYRDLRKGYVVPYTGKWEGBLGTDLVSIIPHGRNVYRANIAATE 120  
DB 106 hpflhryqrolstsyrdlrkgyvvpypqgkwegelgtdlvsiiphgrnvtraniaate 165  
QY 121 SDFEFGNSNNEGILGLAVAEIARPDLSLEPFDSLVKQTHVPMLESLQCGAGFPLNQS 180  
DB 166 sdfefngnsnnegilglavaeiarpdlslepfdslvkqthvpmlefsiqcgafplnqs 225  
QY 181 EYLVASVGSMTIIGTIDHSLYTGSMTYTPIRREMYEVIIVREINIGODLKMDCKEYNDK 240  
DB 226 eylvavgsgmtiigtidhslytgsmtypirrewyeviiivreinigodlkmdeckeyndk 285  
QY 241 SIVDSGTTNLRPLPKYFPEAAVKSIRKAASSTKFPDGFMLGEOLVCMQAGTPMNIPIVIS 300  
DB 286 sivdsgttnlrlpkkyfpeaavksirkaasstekfpdgfmlgvolcmwagtpmniipvis 345  
QY 301 LYLMGEVTNQSFRTITLPPQYLRPVEDVATSDDDCYKFAISQSSGTVMGAVIMEGFYV 360  
DB 346 lylmgevtnqsfritllppqylrpvedvatsddcykfaissgstvmgavimegyfvv 405  
QY 361 FDRARRIGFAVSACHVHDEFRTAAVEGPFVLDMEDCGYNIPQDESLMTIAYVMAI 420  
DB 406 fdrarrigfavsachvhdefrtaavegpfvlmedcgylnipqdeslmtiayymai 465  
QY 421 CALFMLPLCLMVCQWRCLRCLROOHDPADDISILK 456  
DB 466 calfmplclmwcqrclrcrlqghddfaddisilk 501

RESULT 5  
AAB07896  
ID AAB07896 standard; Protein; 501 AA.  
XX  
AC AAB07896;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE Amino acid sequence of a human beta-secretase enzyme.  
XX  
KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
XX  
PN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PE 10-FEB-2000; 2000MO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
XX  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
DR N-PSDB; AAs59550, AAs59551.  
XX  
XX

PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
PS Claim 17; Fig 2A; 121pp; English.

CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme.

SO Sequence 501 AA;

Query Match 100.0%; Score 2419; DB 21; Length 501;  
Best Local Similarity 100.0%; Pred. No. 4.5e-242;  
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSPFVEMVDNLKRGSGGYVEMTVGSPPTLNIIVDTGSSNFVAGAAP 60  
DB 46 etdeepeepgrgsfvevmdnlrksggyvemtvgspptlnllydtgssnfavgaap 105  
QY 61 HPFLHRYQROLSTSYRDLRKGYVVPYTGKWEGBLGTDLVSIIPHGRNVYRANIAATE 120  
DB 106 hpflhryqrolstsyrdlrkgyvvpypqgkwegelgtdlvsiiphgrnvtraniaate 165  
QY 121 SDFEFGNSNNEGILGLAVAEIARPDLSLEPFDSLVKQTHVPMLESLQCGAGFPLNQS 180  
DB 166 sdfefngnsnnegilglavaeiarpdlslepfdslvkqthvpmlefsiqcgafplnqs 225  
QY 181 EYLVASVGSMTIIGTIDHSLYTGSMTYTPIRREMYEVIIVREINIGODLKMDCKEYNDK 240  
DB 226 eylvavgsgmtiigtidhslytgsmtypirrewyeviiivreinigodlkmdeckeyndk 285  
QY 241 SIVDSGTTNLRPLPKYFPEAAVKSIRKAASSTKFPDGFMLGEOLVCMQAGTPMNIPIVIS 300  
DB 286 sivdsgttnlrlpkkyfpeaavksirkaasstekfpdgfmlgvolcmwagtpmniipvis 345  
QY 301 LYLMGEVTNQSFRTITLPPQYLRPVEDVATSDDDCYKFAISQSSGTVMGAVIMEGFYV 360  
DB 346 lylmgevtnqsfritllppqylrpvedvatsddcykfaissgstvmgavimegyfvv 405  
QY 361 FDRARRIGFAVSACHVHDEFRTAAVEGPFVLDMEDCGYNIPQDESLMTIAYVMAI 420  
DB 406 fdrarrigfavsachvhdefrtaavegpfvlmedcgylnipqdeslmtiayymai 465  
QY 421 CALFMLPLCLMVCQWRCLRCLROOHDPADDISILK 456  
DB 466 calfmplclmwcqrclrcrlqghddfaddisilk 501

RESULT 6  
AAB6573  
ID AAB6573 standard; Protein; 503 AA.  
XX  
AC AAB6573;  
XX  
DT 12-APR-2001 (first entry)  
XX  
DE Human pro-memapsin 2.  
XX  
KW Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;



KW APP: memapsin 2 inhibitor; Alzheimer's disease; ss.  
 XX Homo sapiens.  
 OS WO200100665-A2.  
 PN 04-JAN-2001.  
 XX 27-JUN-2000; 2000WO-US17742.  
 PF 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA (UNIT ) UNIV ILLINOIS FOUND.  
 XX Tang JUN, Hong L, Ghosh AK;  
 PI WPI; 2001-137933/14.  
 DR  
 XX  
 XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 XX  
 XX Example 4; Fig 1; 86pp; English.  
 PS  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 XX Sequence 503 AA:  
 SO  
 Query Match 100.0%; Score 2419; DB 22; Length 503;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-242;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ETDEEPEEPGRGSEVENVNLRGSGGGYVEMTVGSPQTLNILDVGTSSNFAVGAAP 60  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC etdeepeepgrgsfvenmdnlrgsgggyvemtvgspqtlnlldtgsnfavgaap 107  
 QY 61 HPELHRYTORLSSTYRDARKGVYVYTOGKMEGELGDLVSIHPGNVTVRANIAATTE 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC hpeflhrytorlsstyrdrarkgvvyvpytqgkwegelgdlvsihpgnvtvranaaite 167  
 QY 108 HPELHRYTORLSSTYRDARKGVYVYTOGKMEGELGDLVSIHPGNVTVRANIAATTE 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC hpeflhrytorlsstyrdrarkgvvyvpytqgkwegelgdlvsihpgnvtvranaaite 167  
 QY 121 SSKFFINGSNMEGILGLAVAEIARPDLSLEPFDSLVRKOTHPNLFSLQLGAGFPLNQS 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC sskffingsnwegilglayaeiarpdsllepffdsivkqthvnpnlfslqcgagfplngs 227  
 QY 181 EYLAIVGSGSMITIGIDHSLYTGSLWYTPIRREMYEVIIVVEINGQDLKMDCKEYNYDK 240  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC eylvavsgsmilgglidhsllytgslywtpirrewyevliivveingqdlkmdckeynydk 287  
 QY 241 STVDSGTNLRPLPKVFEAAVKAASSTKFPDGFMLGSLQCMQAGTTPMNIIFPVIS 300  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC stvdsqgtcnlrlpkvfeaaavksikaasstekfpdgflwgeqlvcwqagtpmnilfpvis 347  
 QY 301 LYLMGEVYNOSFRITILEQQLRYVEDVAISQDDCYKFAISQSSSTGTVMGAVIMEGFYV 360  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC lylmgevynogfrtilleqqlyrvedvatsqddcykfaisqssstgtvmgavimegfyyv 407  
 QY 361 FDRARKRIGFVAVSACHVDEFRTAAVGPFYTLMEDCGYNIPTDSTLMTIAYVMAAI 420  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC fdrarkrigrfavsachvdefrtaavgpfytlmedcgyniprtdstlmtiayvmaai 467

DB 408 fdrarkrigrfavsachvdefrtaavgpfytlmedcgyniprtdstlmtiayvmaai 467  
 QY 421 CALFMLPLCLMWCOMRCLRCLRQOHDPADFDADSLIK 456  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC calfmplclmwcomrcrlcrlrqohdffdaddslilk 503  
 DB 468 calfmplclmwcomrcrlcrlrqohdffdaddslilk 503  
 RESULT 7  
 AAB61335  
 ID AAB61335 standard; protein; 503 AA.  
 XX  
 AC AAB61335;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE T7 promoter and vector sequence.  
 XX  
 KW Memapsin 2; catalyst; Alzheimer's.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200100663-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-US17661.  
 XX  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA  
 PI Tang JUN, Ian X, Koelsch G;  
 XX  
 DR WPI; 2001-102885/11.  
 XX  
 XX Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 XX  
 PS Disclosure; Fig 1; 86pp; English.  
 XX  
 CC The present invention relates to a purified recombinant  
 CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.  
 CC  
 XX Sequence 503 AA:  
 SO  
 Query Match 100.0%; Score 2419; DB 22; Length 503;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-242;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ETDEEPEEPGRGSEVENVNLRGSGGGYVEMTVGSPQTLNILDVGTSSNFAVGAAP 60  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC etdeepeepgrgsfvenmdnlrgsgggyvemtvgspqtlnlldtgsnfavgaap 107  
 QY 61 HPELHRYTORLSSTYRDARKGVYVYTOGKMEGELGDLVSIHPGNVTVRANIAATTE 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC hpeflhrytorlsstyrdrarkgvvyvpytqgkwegelgdlvsihpgnvtvranaaite 167  
 QY 108 HPELHRYTORLSSTYRDARKGVYVYTOGKMEGELGDLVSIHPGNVTVRANIAATTE 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC hpeflhrytorlsstyrdrarkgvvyvpytqgkwegelgdlvsihpgnvtvranaaite 167  
 QY 121 SSKFFINGSNMEGILGLAVAEIARPDLSLEPFDSLVRKOTHPNLFSLQLGAGFPLNQS 180  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC sskffingsnwegilglayaeiarpdsllepffdsivkqthvnpnlfslqcgagfplngs 227  
 QY 168 sskffingsnwegilglayaeiarpdsllepffdsivkqthvnpnlfslqcgagfplngs 227  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC sskffingsnwegilglayaeiarpdsllepffdsivkqthvnpnlfslqcgagfplngs 227  
 QY 181 EYLAIVGSGSMITIGIDHSLYTGSLWYTPIRREMYEVIIVVEINGQDLKMDCKEYNYDK 240  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC eylvavsgsmilgglidhsllytgslywtpirrewyevliivveingqdlkmdckeynydk 240

```

|||||
Db 228 EVLASVGGSMIIGIDHSLYGSILWYPIRREWEYEVLIIVREINGDLMKDCKEYNDK 287
QY 241 SIYDGTNLRLPKRVEAAVKSITKAASSTFKFPDGFNLGSLYCWQAGTTPNNIPVIS 300
Db 248 SIVDSGTNLRLPKKVFEEAAVKSITKAASSTFKFPDGFNLGSLYCWQAGTTPNNIPVIS 347
QY 301 LYIMGEVYNOSFRTIILPQOYLRPEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 348 IYIMGEVYNOSFRTIILPQOYLRPEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 407
QY 361 FDRARRKRGFAVSACHVDEFTAAVEGPEVTLDMEDCGYVNPOTDESTLMTIAYMAAI 420
Db 408 FDRARRKRGFAVSACHVDEFTAAVEGPEVTLDMEDCGYVNPOTDESTLMTIAYMAAI 467
QY 421 CALFMLPCLMWCQRCRLRQOHDHDFADDISILK 456
Db 468 CALFMLPCLMWCQRCRLRQOHDHDFADDISILK 503

RESULT 8
AAE10629
ID AAE10629 standard; Protein: 501 AA.
AC AAY88425;
XX AAY88425;
XX 03-AUG-2000 (first entry)
DT Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
DE Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
XX Alzheimer's disease; beta secretase site.
XX Homo sapiens.
XX WO200017369-A2.
XX 30-MAR-2000.
XX 23-SEP-1999; 99WO-US20881.
XX 24-SEP-1998; 98US-0101594.
XX (PHAA) PHARMACIA & UPJOHN CO.
XX Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Van R;
XX WPI: 2000-303209/26.
XX N-PSDB; AAA15662.
XX PT New enzyme designated human aspartase useful in research into
XX Alzheimer's Disease is capable of cleaving amyloid protein precursor at
XX the beta secretase site to produce amyloid beta peptide.
XX PS Claim 48; Fig 2; 183pp; English.
XX
XX This sequence represents the human aspartyl protease 2 (Asp2) amino acid
XX sequence. The invention relates to a protease (e.g. Asp2) capable of
XX cleaving the beta secretase site of amyloid precursor protein (APP). The
XX sequence contains a sequence encoding the amino acid sequence DTG and a
XX sequence encoding DSG or DTG separated by 100-300 amino acids. When
XX mutated the APP gene causes an autosomal dominant form of Alzheimer's
XX disease. APP localises to the cell surface membrane and have a single
XX C-terminal transmembrane domain. Proteolytic processing of APP produces
XX the amyloid beta protein, which is possibly very important in Alzheimer's
XX disease. The invention includes a nucleotide sequence encoding the
XX protease, a vector containing the nucleotide sequence, and a cell line
XX comprising the vector. Methods for screening for inhibitors of beta
XX secretase activity are also given in the invention. The human aspartase
XX protein and nucleotide sequences and the methods for identifying
XX inhibitors of the protease, are useful in the treatment of and research
XX in to Alzheimer's disease.
XX

```

```

SQ Sequence 501 AA:
Query Match 99.8%; Score 2414; DB 21; Length 501;
Best Local Similarity 99.8%; Pred. No. 1,5e-241;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ETDEEPEERGRSGSFYEMVDNLRGSGGGYVEMTQSPROTINILVDPGSSNFAYGAAP 60
Db 46 EIDEEPEERGRSGSFYEMVDNLRGSGGGYVEMTQSPROTINILVDPGSSNFAYGAAP 105
QY 61 HPFLHRYQROLSSSTYRDLKRGYVPTTQKMEGELGTDLVSIHPGPNVTNANIAITE 120
Db 106 HPFLHRYQGLASTYLDLKGYVPTTQKMEGELGTDLVSIHPGPNVTNANIAITE 165
QY 121 SDRFFINGSMWEGILGLAYEIAIARPDSDLFPFDLSLYKQTHVNNLSIOLCGAFPLNOS 180
Db 166 SDRFFINGSMWEGILGLAYEIAIARPDSDLFPFDLSLYKQTHVNNLSIOLCGAFPLNOS 225
QY 181 EYLASVGGSMIIGIDHSLYGSILWYPIRREWEYEVLIIVREINGDLMKDCKEYNDK 240
Db 226 EVLASVGGSMIIGIDHSLYGSILWYPIRREWEYEVLIIVREINGDLMKDCKEYNDK 285
QY 241 SIYDGTNLRLPKRVEAAVKSITKAASSTFKFPDGFNLGSLYCWQAGTTPNNIPVIS 300
Db 286 SIVDSGTNLRLPKKVFEEAAVKSITKAASSTFKFPDGFNLGSLYCWQAGTTPNNIPVIS 345
QY 301 LYIMGEVYNOSFRTIILPQOYLRPEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 360
Db 346 IYIMGEVYNOSFRTIILPQOYLRPEDVATSDDCYKFAISQSSTGTVMGAVIMEGFYV 405
QY 361 FDRARRKRGFAVSACHVDEFTAAVEGPEVTLDMEDCGYVNPOTDESTLMTIAYMAAI 420
Db 406 FDRARRKRGFAVSACHVDEFTAAVEGPEVTLDMEDCGYVNPOTDESTLMTIAYMAAI 465
QY 421 CALFMLPCLMWCQRCRLRQOHDHDFADDISILK 456
Db 466 CALFMLPCLMWCQRCRLRQOHDHDFADDISILK 501

RESULT 9
AAE10629
ID AAE10629 standard; Protein: 501 AA.
AC AAE10629;
XX 10-DEC-2001 (first entry)
DT Human aspartyl protease 2(a) [hu-Asp2(a)] protein.
DE Human aspartyl protease 2(a) [hu-Asp2(a)] protein.
XX Human; aspartyl protease 2(a); Asp2(a); amyloid precursor protein; APP;
XX Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
XX amyloid plaque; neuronal loss; proteolytic; neurotrophic; neuroprotective;
XX chromosome 11q23.3-24.1.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX Peptide /label= Signal_peptide
XX Peptide 22..45
XX Peptide /label= Asp_2a_prepropeptide
XX Peptide 46..57
XX Peptide /label= Asp_2a_propeptide
XX Protein 58..501
XX Protein /label= Mature_human_Asp_2a_protein
XX Region 420..454
XX Region /label= Alpha_helical_spacer_region
XX Domain 455..477
XX Domain /label= Transmembrane_domain
XX Domain 478..501
XX Domain /label= Cytoplasmic_domain
XX

```

PN GB235767-A.  
XX  
PD 04-JUL-2001.  
XX  
PF 22-SEP-2000; 2000GB-0023315.  
XX  
PR 23-SEP-1999; 99US-0155493.  
PR 23-SEP-1999; 99US-0404133.  
PR 23-SEP-1999; 99WO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Bienkowski MJ, Gurney M;  
XX  
DR WPI: 2001-444208/48.  
DR N-PSDB; AAD17865.  
XX  
XX  
PT Polypeptide comprising fragments of human aspartyl protease with  
PT amyloid precursor protein processing activity and alpha-secretase  
PT activity, for identifying modulators useful in treating Alzheimer's  
PT disease -  
XX  
XX  
PS Example 2; Fig 2; 187pp; English.

CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified  
CC Asp1 proteins which lack transmembrane domain or amino terminal  
CC domain or cytoplasmic domain and retains alpha-secretase activity  
CC and amyloid protein precursor (APP) processing activity. The proteins  
CC of the invention are useful for assaying hu-Asp1 alpha-secretase  
CC activity, which in turn is useful for identifying modulators of  
CC hu-Asp1 alpha-secretase activity, where modulators that increase  
CC hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's  
CC disease (AD) which causes progressive dementia with consequent  
CC formation of amyloid plaques, neurofibrillary tangles, gliosis and  
CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying  
CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with  
CC the substrate under acidic conditions and determining the level of  
CC hu-Asp1 proteolytic activity. The present sequence is long form of  
CC human Asp2 protein, designated as Asp2(a). Asp2 gene is localised  
CC on chromosome 11q23.3-24.1.  
XX  
XX  
SQ Sequence 501 AA:

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EFDDEPEEPEGRGSGFVEMVDNRKSGSGGYVEMTVGSEPPQTNLILVDTGSSNFAVGAAP 60  
DB 46 EEDDEPEEPEGRGSGFVEMVDNRKSGSGGYVEMTVGSEPPQTNLILVDTGSSNFAVGAAP 105  
OY 61 HPELHRYTORQLSTYRDLRKGVYVYTOGKEGELGDIVSIPHGPRVTVRANAATTE 120  
DB 106 hpfllhryygrqlsstyrldlrkyvvypryqgkwegeigcdlvsipgprvtyranaahte 165  
OY 121 SDKFTINGSNNEGIIGLAVAEIARPDLSLEPPFDLSLVKQTHVPLNLSQLGAGFPLNQ 180  
DB 166 sckffingsnwegllglaaeiaarpdlslepffdsllvqthvplnfsihlgagfplnqs 225  
OY 181 EFLASVGSMTIIGCDHLYTGSIMWYPIREKMYEVIIVAEINSGDLKMDCKEYVNDK 240  
DB 226 evlasvgsmtiigcdhlytgsimwypirrekyeviiivaeinsgdlkmdckeyvndk 285  
OY 241 SIVDSGTTLNLPRKVFEEAAVKSIAKASTKFPDGFMLGEOLVCMOGCTPMNIEPITS 300  
DB 286 sivdsgttlnlprkvfeaaavksikaastekfpdgfmlgeqlvcwagqtlcpmniifpvis 345  
OY 301 LYLWGEVYNOSFRITILPQOYLREVEDVATISQDDCYKFAISQSSGTGVMAVINEGFYV 360  
DB 346 lylwgevynosfrtilpqyrlrpevdatsqddcykfaisqssgtgvmaivegfyyv 405

OY 361 FDRARRIGFANVSACHVHDEFFRTAAVEGPFVTLMDGCGYNIPQDESTLMTIAYVMAI 420  
DB 406 fdrarrigfavsachvhdefrtaaevpfvclmdedcgnipqdestlmtiayvmai 465  
OY 421 CALFMDPLCIAMWCORCLRCRQOHDDFADDISLKL 456  
DB 466 calfmplclmwcqrcrlcrlrqhddadidislkl 501

RESULT 10  
AAE06859  
ID AAE06859 standard; Protein; 501 AA.  
XX  
AC AAE06859;  
XX  
DT 23-OCT-2001 (first entry)  
XX  
DE Human aspartyl protease 2a (Hu-Asp2a) protein.  
XX  
KM Human; aspartyl protease 2a; Asp 2a; beta-amyloid precursor protein; APP;  
KM beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;  
KM neurofibrillary tangle; neuronal loss; amyloid-beta peptide; neurotropic;  
KM neuroprotective; antisense therapy; gene therapy;  
KM chromosome 11q23.3-24.1.  
XX  
OS Homo sapiens.

XX  
XX  
FH Key Location/Qualifiers  
FH Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..501  
FT Region /note= "Mature human aspartyl protease 2a (Hu-Asp2a)"  
FT Domain /note= "Alpha helical spacer region"  
FT Domain /label= "Transmembrane\_domain  
FT Domain /label= "Transmembrane\_domain  
FT Domain /label= "Transmembrane\_domain  
FT Domain /label= "Transmembrane\_domain  
XX  
XX  
XX WO200150829-A2.  
XX  
XX 19-JUL-2001.  
XX  
XX 09-MAY-2001; 2001WO-IB00799.  
XX  
XX 09-MAY-2001; 2001WO-IB00799.  
XX  
XX  
XX (BIEN/) BIENKOWSKI M J.  
XX (GURN/) GURNEY M E.  
XX (HEIN/) HEINRIKSON R L.  
XX (PARO/) PARODI L A.  
XX (YANR/) YAN R.

XX  
XX Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
XX  
XX WPI: 2001-483072/52.  
XX  
XX N-PSDB; AAD13021.  
XX  
XX  
XX Novel purified polypeptide comprising fragment of mammalian aspartyl  
XX protease 2, lacking Asp2 transmembrane domain and retaining beta  
XX secretase activity of Asp2 useful for identifying inhibitors of Asp2  
XX activity -  
XX  
XX Claim 49; Fig 2; 185pp; English.

XX  
XX The invention relates to human aspartyl proteases (Hu-Asp), beta-amyloid  
XX precursor protein (APP) isoforms and their corresponding DNA molecules.  
XX Human aspartyl proteases can act as beta-secretase proteases useful for  
XX treating Alzheimer's disease. APP isoforms are useful for identifying  
XX modulators of amyloid-beta peptide production, for use in designing  
XX therapeutics for the treatment and prevention of Alzheimer's disease,  
XX dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis

CC and neuronal loss. APP isoforms are also used in methods for identifying  
 CC inhibitors and modulators of human Asp2 activity. The invention relates  
 CC to a method for identifying agents that modulate the activity of human  
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
 CC as a means to screen in cellular assays for the inhibitors of beta- and  
 CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
 CC polymerase chain reactions (PCR). The probes are useful for detecting  
 CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
 CC blots. The present sequence is human aspartyl protease 2 (Hu-Asp2), a  
 CC 'long' form designated as (Hu-Asp2a). Hu-Asp 2 gene is localised on  
 CC chromosome 11q23.3-24.1.

XX Sequence 501 AA:

Query Match 99.8%; Score 2414; DB 22; Length 501;  
 Best local Similarity 99.8%; Pred. No. 1.5e-241;  
 Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGSGGYVEMTVGSPQTLNLTVDGSSNFAVGAP 60  
 |||||||  
 DB 46 etdepeepgrgsgfivemvndlrksgsggyvemtvgspqtlnlvdtgssnfavgap 105  
 |||||||  
 QY 61 HPELHRYGROLSSRYDLRKGVVYPYTOGKMEGELGTDLVSIIPHGPNTVVRANIAATTE 120  
 |||||||  
 DB 106 hpeflhryyrolsstyrdlrkgyvypytogkwegelgtldvsiipghpntvvranaiaate 165  
 |||||||  
 QY 121 SDKFFINGSNMGGIIGLAVAEIARPDLSLEPPDLSIKOTHPNLFSLDLCAGAPFLNOS 180  
 |||||||  
 DB 166 sdkffingsnmgwiiglayaeiarpdlslepfdsivkqthvnlfsllcqagfplnds 225  
 |||||||  
 QY 181 EYIASVSGSMIIIGIDHSIYTGSLWYTPIRBMYEVIIVAREINQODKMKCKEKNYDK 240  
 |||||||  
 DB 226 eviasvsgsmliigidhslytgsliwtpirreyevliivareinqqdkmkckeynyk 285  
 |||||||  
 QY 241 SYDSGTTNLRLPKRYFEAAVKSIRKASSTKRRPDGFWLCEOLVCMQAGTPWNIPVTS 300  
 |||||||  
 DB 286 sydsgstnlrlpkryfeaaavksirkaasstekrpdgfwlgeqivcwaqgtlpwnlfpvts 345  
 |||||||  
 QY 301 LYTMGEVNTOSRITITLPOQYLRPEVDVATSQDDCYKFAISQSTGTVMGAVMEGFYVV 360  
 |||||||  
 DB 346 lytmgevtngsfrtlllpqylyrpevdvatsqddcykfaisgstgtvmgavmeqfyvv 405  
 |||||||  
 QY 361 FDMARRIGAVASACHVHDEFRTAAVEGPFVTLDMDCGYNIPQTESTLMTIAYMAAI 420  
 |||||||  
 DB 406 fdmartrigavasachvdeftaaavegpfvtldmedcgynipqtdestlmtlaymaai 465  
 |||||||  
 QY 421 CALFMLPLCLMVCWMRCLRCLROQHDDFADDISILK 456  
 |||||||  
 DB 466 calfmplclmvcwrcrlcrlrqhddfaddisilk 501  
 |||||||

RESULT 11

AAU06603 standard: Protein; 501 AA.

XX AAU06603:

XX 24-OCT-2001 (first entry)

XX Human Aspartyl protease 2(a), Asp2(a).

XX Human: Aspartyl protease; Asp2(a); beta-secretase; neurotrophic;

XX neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;

XX amyloid-beta; Abeta.

XX Homo sapiens.

XX Key location/Qualifiers

FT Peptide 1..21  
 FT /label= signal peptide  
 FT 22..45  
 FT /label= pre\_pro-peptide  
 FT Peptide

FT Peptide 46..57  
 FT /label= Pro-peptide  
 FT Protein 57..501  
 FT /label= Mature\_Asp2(a)  
 FT Region 420..454  
 FT /label= Alpha\_helical\_spacer\_region  
 FT Domain 455..477  
 FT /label= Transmembrane\_domain  
 FT Domain 478..501  
 FT /label= Cytoplasmic\_domain

PN W0200149098-A2.

PD 12-JUL-2001.

XX 09-MAY-2001; 2001WO-IB00798.

XX 09-MAY-2001; 2001WO-IB00798.

PA (BIEN/) BIENKOWSKI M J.

PA (GURN/) GURNEY M E.

PA (HEIN/) HEINRIKSON R L.

PA (PARO/) PARODI L A.

PA (YANR/) YAN R.

PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;

XX WPI: 2001-502549/55.

DR N-PSDB; AAS11517.

PS Claim 49; Fig 2; 185pp; English.

The invention relates to a purified polypeptide comprising a fragment of  
 CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2  
 CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
 CC the fragment retain the beta-secretase activity of the mammalian Asp2  
 CC protein. The invention also details polynucleotides for the Asp  
 CC proteins and vectors expressing them, and a polypeptide (isoform of  
 CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
 CC APP or its fragment containing an APP cleavage site recognizable by a  
 CC mammalian beta-secretase, and further comprising two lysine residues at  
 CC the carboxyl terminus of the amino acid sequence are methods of identifying  
 CC APP fragment. Also included in the invention are methods of identifying  
 CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
 CC useful for treating Alzheimer's disease. APP is useful in methods for  
 CC identifying inhibitors or modulators of human Asp2 activity and  
 CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
 CC therapeutics for the treatment or prevention of Alzheimer's disease.  
 CC APP comprising the APP-SW-beta-secretase peptide sequence (NDA), which  
 CC is associated with increased levels of Abeta processing is useful in  
 CC assays relating the Alzheimer's research. The expression vector is useful  
 CC for recombinantly expressing APP. Nucleic acids that hybridise to  
 CC Asp oligonucleotides are useful as probes or primers. The probes are  
 CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
 CC Northern and Southern blots. The present sequence is human Asp2(a).

XX Sequence 501 AA:

Query Match 99.8%; Score 2414; DB 22; Length 501;  
 Best local Similarity 99.8%; Pred. No. 1.5e-241;  
 Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGSGGYVEMTVGSPQTLNLTVDGSSNFAVGAP 60  
 |||||||  
 DB 46 etdepeepgrgsgfivemvndlrksgsggyvemtvgspqtlnlvdtgssnfavgap 105  
 |||||||  
 QY 61 HPELHRYGROLSSRYDLRKGVVYPYTOGKMEGELGTDLVSIIPHGPNTVVRANIAATTE 120  
 |||||||

Db 106 hpfllhyrqlsstyrdrirkgyvpytgkgwgelgldslsiphgpnvtraniaalte 165  
QY 121 SDFEFGNSNMGITGLAAEAIAEPDDELPEFPDLSLVQTHVNLFSLOLCGAGFPLNQS 180  
Db 166 sckffingnwegllgldhslayaelarppdslepfidslvkqthvnlfsllhlcgagfplnqs 225  
QY 181 EYLASVGSMTIGGIDHSIYTGSLWYTPIRREMYEVIIVREINGODLKMCKEYNDK 240  
Db 226 evlasvgsmligldhslayaelarppdslepfidslvkqthvnlfsllhlcgagfplnqs 285  
QY 241 SIYDSGTTNLRPKKVEFAAKSIRKAASSTEFDPDGEVLGQVLCWQAGTTPANNIFPVIS 300  
Db 286 slvdsqgtlnlrpkkvfaavksikaasstekfpdgfwlgeqlvcwgagtcplnlfpyis 345  
QY 301 LYLMGEVTVNSFRITILPQOYLRPVEDVATSDDCYKFAISQSGTGVMAVIMEGFYV 360  
Db 346 lylmgevtnsfriltpqyylrpvedvatsqddcykfalsqsgtgvmaavimegfyvv 405  
QY 361 FDRARRKIGFAVSACHVDEFFRTAAVEGPFVTLDMDCGYNIPTQDESTLMTIAYVMAAI 420  
Db 406 fdrarrkigfavsachvdeffrtaaavegpfvclmedcgynipqtdestlmtlayymaa 465  
QY 421 CALFMLPLCIAMVCOMRCRLRQOHDDFADDSILK 456  
Db 466 calfmplclmwcgvrcrlrcrlqghddfaddisllk 501

RESULT 12

AAU07202 standard; Protein; 501 AA.

AAU07202:

24-OCT-2001 (first entry)

Human aspartyl protease 2a (Asp-2a).

Human: aspartyl protease 1; Asp-1; nootropic; neuroprotective;

Aspartyl protease 2; Asp2; amyloid protein precursor; App;

beta-secretase; Alzheimer's disease.

Homo sapiens.

Key Location/Qualifiers

FT Peptide 1..21

FT MISC\_feature 22..45

FT MISC\_feature 46..57

FT Protein 58..501

FT Region 420..454

FT Domain 455-477

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

FT Domain 478..501

XX  
PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;  
XX  
XX WPI: 2001-502548/55.  
DR N-PSDB; AAS11702.  
XX  
PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
PT activity  
XX  
PS Claim 49; Fig 2; 185pp; English.  
XX  
CC The invention relates to a novel purified polypeptide comprising a  
CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the  
CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide  
CC and the fragment retain the beta-secretase activity of the mammalian Asp2  
CC protein. Also included is an isoform of amyloid protein precursor (APP)  
CC comprising the amino acid sequence of a APP or its fragment containing  
CC an APP cleavage site recognizable by a mammalian beta-secretase, and  
CC further comprising two lysine residues at the carboxyl terminus of the  
CC amino acid sequence of the mammalian APP or APP fragment. The  
CC polypeptides are used for assaying for modulators of beta-secretase  
CC activity; identifying agents that inhibit the APP processing activity  
CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that  
CC modulate the activity of Asp2; and for reducing cellular production of  
CC amyloid beta (A-beta) from APP. Agents identified by the above methods  
CC are useful for treating Alzheimer's disease; and for identifying  
CC modulators of amyloid-beta (beta) peptide production, for use in  
CC designing therapeutics for the treatment or prevention of Alzheimer's  
CC disease. Probes and primers derived from Asp nucleic acid sequences  
CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
CC Northern and Southern blots. The present sequence represents the  
CC amino acid sequence of human Asp-2a used in the methods of the invention.  
XX  
SQ Sequence 501 AA;

Query Match 99.88; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.88; Pred. No. 1.5e-241;

Matches 453; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSEFVENVNLIRKSGQYVEMTVSSPOTINIIVDSSNFANVGAAP 60  
Db 46 etdepeepgrgsfvenvnlirgksqgyvemyvsgpqtlnllvdygssnfavgaap 105  
QY 61 HPEIHRYYQRLSTYDLKRGYVVPYTGKGWEGELGTDLVSIPIHGPNTVVRANIAITE 120  
Db 106 hpeihryyqrlstydldrkgvypvtgkgwgelgtldvsiiphgpnvtraniaalte 165  
QY 121 SDFEFGNSNMGITGLAAEAIAEPDDELPEFPDLSLVQTHVNLFSLOLCGAGFPLNQS 180  
Db 166 sckffingnwegllgldhslayaelarppdslepfidslvkqthvnlfsllhlcgagfplnqs 225  
QY 181 EYLASVGSMTIGGIDHSIYTGSLWYTPIRREMYEVIIVREINGODLKMCKEYNDK 240  
Db 226 evlasvgsmligldhslayaelarppdslepfidslvkqthvnlfsllhlcgagfplnqs 285  
QY 241 SIYDSGTTNLRPKKVEFAAKSIRKAASSTEFDPDGEVLGQVLCWQAGTTPANNIFPVIS 300  
Db 286 slvdsqgtlnlrpkkvfaavksikaasstekfpdgfwlgeqlvcwgagtcplnlfpyis 345  
QY 301 LYLMGEVTVNSFRITILPQOYLRPVEDVATSDDCYKFAISQSGTGVMAVIMEGFYV 360  
Db 346 lylmgevtnsfriltpqyylrpvedvatsqddcykfalsqsgtgvmaavimegfyvv 405  
QY 361 FDRARRKIGFAVSACHVDEFFRTAAVEGPFVTLDMDCGYNIPTQDESTLMTIAYVMAAI 420  
Db 406 fdrarrkigfavsachvdeffrtaaavegpfvclmedcgynipqtdestlmtlayymaa 465  
QY 421 CALFMLPLCIAMVCOMRCRLRQOHDDFADDSILK 456  
Db 466 calfmplclmwcgvrcrlrcrlqghddfaddisllk 501

RESULT 13  
AAE02581  
ID AAE02581 standard; Protein; 501 AA.  
XX AAE02581;  
XX  
XX 10-AUG-2001 (first entry)  
XX  
XX  
XX Human aspartyl protease 2a (asp 2a).  
XX  
XX Human: alpha-secretase; amyloid precursor protein; APP, therapy;  
XX Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp 2a;  
XX beta-secretase; chromosome 11q23.3-24.1.  
XX  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..21  
XX /label= Signal\_peptide  
XX Peptide 22..45  
XX /label= Asp\_2a-prepropeptide  
XX Peptide 46..57  
XX /label= Asp\_2a-propeptide  
XX Protein 58..501  
XX /label= Mature\_human\_Asp-2a-protein  
XX Active-site 93..95  
XX /label= Active\_site\_1  
XX Active-site 289..291  
XX /label= Active\_site\_2  
XX Region 420..454  
XX /label= Alpha\_helical\_spacer  
XX Domain 455..477  
XX /label= Transmembrane\_domain  
XX Domain 478..501  
XX /label= Cytoplasmic\_domain  
XX Region 486..501  
XX /note= "Peptide #2"  
XX  
XX  
XX WO200123533-A2.  
XX  
XX  
XX 05-APR-2001.  
XX  
XX  
XX 22-SEP-2000; 2000WO-US26080.  
XX  
XX 23-SEP-1999; 99US-0155493.  
XX 23-SEP-1999; 99WO-US20881.  
XX 13-OCT-1999; 99US-0416901.  
XX 06-DEC-1999; 99US-0169232.  
XX  
XX (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
XX  
XX Gurney M, Bienkowski MJ;  
XX  
XX WPI: 2001-290516/30.  
XX N-PSDB; AAD06739.  
XX  
XX Enzymes that cleave the alpha-secretase site of the amyloid precursor  
XX protein, useful for the treatment of Alzheimer's disease -  
XX  
XX Example 2; Fig 2; 189pp; English.  
XX  
XX The present invention relates to enzymes for cleaving the alpha-  
XX secretase site of the amyloid precursor protein (APP) and methods of  
XX identifying those enzymes. The methods may be used to identify enzymes  
XX that may be used to cleave the alpha-secretase cleavage site of the APP  
XX protein. The enzymes may be used to treat or modulate the progress of  
XX Alzheimer's disease. The present sequence is human aspartyl protease 2a  
XX (Asp 2a). Asp 2a has beta-secretase protease activity. Asp2 gene  
XX is located on chromosome 11q23.3-24.1.  
XX  
XX Sequence 501 AA:  
XX  
XX

Query Match 99.8%; Score 2414; DB 22; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1,5e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 EPTDEEPEEPGRGSGFVEMVNLGKSGGYYVEMTVGSPPTLNLVDTGSSNFAVGAAP 60  
DB 46 etdeepeepgrirgsfivemvnlrgksgggyvemtvgspptclnlivdvgssnfaavgaap 105  
OY 61 HPELHRYQRQLSSTRDLRKGYVYPTTGKMEGEGTDLVSLPHGPNYVRANIAAITE 120  
DB 106 hpfllhryqqrqlsstrydlrkyvypylcgkwegelqdlvslphgpnvtraniaaite 165  
OY 121 SDRFINGSWEGILGATAYEAIARPDSLEPPDSLVKQTHVNLFSLOLGAAPPLMS 180  
DB 166 sdrfingswegilgatayeaiaarpdsleppdslvkqthvnlfsllhcgagfplngs 225  
OY 181 EVLASVGSMLIGIDHSLYTGSLSWTPPIREMYEVYIYVREINGODLKMDCKEYNDK 240  
DB 226 evlasvgsmliigidhslytgsllswtpiremyevilyvreinngodlkmckeynydk 285  
OY 241 STVDSGTTNLRPKVFEAAVKSITKAASSTKEPDPGFWLGEOLVCMQAQTPPNIFPVIS 300  
DB 286 stvdsqtnlrlpkvfeaaavksikaasstekfpgdfwlgelvcwgaqtpnifpvls 345  
OY 301 LYLGEVITNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSTGYMKAVIMEGYYV 360  
DB 346 lylgevtngsfrtilpqpqylrpvedatsqdcykfaisqstgtyvmgavimegylyv 405  
OY 361 FDRARRIRGFVAVSACHVDEFTAAVEGFWTLDMECGYNIPTDSTLMTIAYVMAAT 420  
DB 406 fdrarkrigrfvavschvdeftaaavegfwtlmedcgyinipdstlmtiayvmaat 465  
OY 421 CALFMLPLCLMWQWRCRLRQOHDDFADDISLK 456  
DB 466 calfmplclmwqcwrcrlrqohddfaaddisllk 501  
RESULT 14  
AAM59807  
ID AAM59807 standard; Protein; 501 AA.  
XX  
XX  
XX AAM59807;  
XX  
XX 26-OCT-1998 (first entry)  
XX  
XX  
XX Amino acid sequence of human ASP2 (aspartic protease 2).  
XX  
XX Human: ASP2; aspartic protease 2; agonist; antagonist; immunospecific;  
XX antibody; inhibition; Alzheimer's disease; cancer; proteinase;  
XX prohormone processing.  
XX  
XX Homo sapiens.  
XX  
XX  
XX EP855444-A2.  
XX  
XX 29-JUL-1998.  
XX  
XX 27-JAN-1998; 98EP-0300573.  
XX  
XX 28-JAN-1997; 97GB-0001684.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
XX Chapman CG, Murphy K, Powell DJ, Smith TS;  
XX  
XX WPI: 1998-389809/34.  
XX N-PSDB; AAVA1696.  
XX  
XX New nucleic acid encoding human aspartic protease 2 - used to treat,  
XX prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone  
XX

PT processing  
XX  
PS Claim 1: Page 7; 26pp; English.  
XX  
CC This is the amino acid sequence of the human ASP2 (aspartic protease  
CC family), used in the method of the invention. Agonists and  
CC antagonists for ASP2 immunospecific antibodies are used to treat  
CC conditions requiring increased or decreased activity or expression of  
CC ASP2 respectively. ASP2 is used to treat and diagnose e.g.,  
CC Alzheimer's disease, cancer and pro hormone processing and ASP2 or a  
CC fragment can be used to induce an immune response against the above  
CC conditions.  
XX  
S0 Sequence 501 AA:

Query Match 99.8%; Score 2413; DB 19; Length 501;  
Best Local Similarity 99.8%; Pred. No. 1.9e-241;  
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGSGGYVEMTVGSPPTQNLIVDTGSSNPAVGAAP 60  
DB 46 etdepeepgrgsgfvevmdnlrksqgyvemtvgspptqnlivdtgssnfavgaap 105  
QY 61 HPFLHRYRQRLSTYRDLRKGVVVPYTOGKWESELTDLVSIHPGPNVTVRANIAITE 120  
DB 106 hpfllhryrqrlsttyrdlrrkgyvvpitogkwegeseltldvsihpgpnvtvraniaite 165  
QY 121 SDRFFINGSMWEGILGLAVAEIARPDLSLEPFDSLVKOTHPVNLPSLQCGAGFPLNOS 180  
DB 166 sdrffingsmwegilglavaeiarpdlslepfdslvkothpvnlpnlslqcgagfplnos 225  
QY 181 EVLASVGSMTIGIDHSILYTGSLMTPIRREYVEIIVRVEINQDLMCKEYNDK 240  
DB 226 evlasvgsmtigidhsilgtslmtpirreyveiiivrveinqdldmckeynydk 285  
QY 241 SIYDSGTTNLRPKKVEFAVKSIRKASTREKPPDGFVGEOLVQMOAGTTPMNFPIVIS 300  
DB 286 siydsgttnlrpkkvefaavksirkaastrekppdgfvgeolvqmoagtppmnpfivis 345  
QY 301 LYLMGEVTNOSRITILPQOYLRPVEDVATSDDDCYKFAISQSTGTVGAVIMEGFYV 360  
DB 346 lylmgevtngsritilppqylrpvedvatsddcykfaisgstgtvmgavimegfyv 405  
QY 361 FDRARRIGFAVSAHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESLMTIATVMAAI 420  
DB 406 fdrarrigfavsachvhdefrtaaavgpfvtldmedcgylnipqdeslmtiatvymaa 465  
QY 421 CALFMRPLCLAMCQMRCLRQOHDDFADDISLAK 456  
DB 466 calfmrlplclamcqmrclrqohddfaddisllk 501

RESULT 15  
AA94769  
ID AA94769 standard; Protein; 501 AA.  
XX  
AC AA94769;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Rat beta-secretase protein.  
XX  
KM Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease; rat;  
KM Down's syndrome; amyloid angiodopathy; gene therapy; neuroprotective; ss.  
OS Rattus sp.  
XX  
PN WO200058479-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 23-MAR-2000; 2000WO-US07755.

XX  
PR 26-MAR-1999; 99US-0277229.  
XX  
PA (AMGE-) AMGEN INC.  
PI Clifton M, Vassar RJ, Bennett BD;  
XX  
XX WPI: 2000-594643/56.  
DR N-PSDB; AAA28280.  
XX  
PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful  
PT for diagnosis and gene therapy of Alzheimer's disease -  
PS  
XX Claim 1: Fig 6; 145pp; English.

This invention relates to 3 nucleotide sequences encoding beta-secretase  
CC proteins. Beta-secretase is an enzyme involved in the production of one  
CC of the components of amyloid plaques involved in Alzheimer's disease. The  
CC invention includes an expression vector comprising the nucleotide  
CC sequence, a host cell comprising the expression vector, and a process for  
CC producing the protein through culturing the transformed cells. Also  
CC included in the invention are a polypeptide derivative of the  
CC beta-secretase protein, a fusion protein comprising beta-secretase fused  
CC to a heterologous amino acid sequence, and a method for modulating the  
CC levels of beta-secretase polypeptide in a mammal comprising administering  
CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and  
CC neurotropic activity. The beta-secretase gene and related genes on chromosomes  
CC and as hybridization probes in diagnostic assays to test for the presence  
CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's  
CC syndrome, and amyloid angiodopathy. The nucleotide sequence may also be  
CC used as anti-sense inhibitors of beta-secretase expression. In gene  
CC therapy of Alzheimer's disease, and for the identification of compounds  
CC that modulate beta-secretase activity. Antibodies to the beta-secretase  
CC protein may be used for in vitro and in vivo diagnostic purposes to  
CC detect the presence of beta-secretase polypeptide in a body fluid or cell  
CC sample. The present sequence represents the rat beta-secretase protein.  
S0 Sequence 501 AA:

Query Match 98.9%; Score 2392; DB 21; Length 501;  
Best Local Similarity 98.2%; Pred. No. 2.9e-239;  
Matches 448; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGSGGYVEMTVGSPPTQNLIVDTGSSNPAVGAAP 60  
DB 46 etdepeepgrgsgfvevmdnlrksqgyvemtvgspptqnlivdtgssnfavgaap 105  
QY 61 HPFLHRYRQRLSTYRDLRKGVVVPYTOGKWESELTDLVSIHPGPNVTVRANIAITE 120  
DB 106 hpfllhryrqrlsttyrdlrrkgyvvpitogkwegeseltldvsihpgpnvtvraniaite 165  
QY 121 SDRFFINGSMWEGILGLAVAEIARPDLSLEPFDSLVKOTHPVNLPSLQCGAGFPLNOS 180  
DB 166 sdrffingsmwegilglavaeiarpdlslepfdslvkothpvnlpnlslqcgagfplnos 225  
QY 181 EVLASVGSMTIGIDHSILYTGSLMTPIRREYVEIIVRVEINQDLMCKEYNDK 240  
DB 226 evlasvgsmtigidhsilgtslmtpirreyveiiivrveinqdldmckeynydk 285  
QY 241 SIYDSGTTNLRPKKVEFAVKSIRKASTREKPPDGFVGEOLVQMOAGTTPMNFPIVIS 300  
DB 286 siydsgttnlrpkkvefaavksirkaastrekppdgfvgeolvqmoagtppmnpfivis 345  
QY 301 LYLMGEVTNOSRITILPQOYLRPVEDVATSDDDCYKFAISQSTGTVGAVIMEGFYV 360  
DB 346 lylmgevtngsritilppqylrpvedvatsddcykfaisgstgtvmgavimegfyv 405  
QY 361 FDRARRIGFAVSAHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDESLMTIATVMAAI 420  
DB 406 fdrarrigfavsachvhdefrtaaavgpfvtldmedcgylnipqdeslmtiatvymaa 465

Wed Oct 30 14:11:47 2002

us-09-724-571-43.rag

Page 12

QY	421 CALFMEPLCIMVCCMRCLRCILRQOHDDPADDISLK	456
Db	466 calfmplrcimvcqwrclrlrhnddlaaddisllk	501

Search completed: August 7, 2002, 09:16:14  
Job time: 159 sec



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 0.766585 Seconds

(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-103

Perfect score: 20

Sequence: 1 VKMD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: pIR1:\*  
2: pIR2:\*  
3: pIR3:\*  
4: pIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	33	2 S23094	beta-amyloid prote
2	20	100.0	52	1 RUME	rubredoxin - Megas
3	20	100.0	57	2 E60045	Alzheimer's diseas
4	20	100.0	57	2 E60045	Alzheimer's diseas
5	20	100.0	57	2 G60045	Alzheimer's diseas
6	20	100.0	57	2 G60045	Alzheimer's diseas
7	20	100.0	57	2 D60045	Alzheimer's diseas
8	20	100.0	57	2 A60045	Alzheimer's diseas
9	20	100.0	57	2 B60045	Alzheimer's diseas
10	20	100.0	74	2 F96670	Alzheimer's diseas
11	20	100.0	75	2 A9799	hypothetical prote
12	20	100.0	84	2 P00438	hypothetical prote
13	20	100.0	86	2 D98327	Alzheimer's diseas
14	20	100.0	86	2 A12955	hypothetical prote
15	20	100.0	90	2 AD3625	cell division topo
16	20	100.0	92	2 T12859	cell division inhi
17	20	100.0	93	2 T23016	hypothetical prote
18	20	100.0	94	2 JN0660	hypothetical prote
19	20	100.0	94	2 S32105	heat shock protein
20	20	100.0	103	2 AC1582	chaperonin groes -
21	20	100.0	103	2 A11228	thioredoxin [impor
22	20	100.0	104	2 H95206	thioredoxin [impor
23	20	100.0	104	2 H98071	thioredoxin [impor
24	20	100.0	105	2 PH1526	thioredoxin reduct
25	20	100.0	127	2 H90315	gamma-aminobutylic
26	20	100.0	134	2 G69444	conserved hypothet
27	20	100.0	140	2 A69445	hypothetical prote
28	20	100.0	142	2 E84069	hypothetical prote
29	20	100.0	145	2 F75189	hypothetical prote

30	20	100.0	146	2 C69136	hypothetical prote
31	20	100.0	147	2 T19162	hypothetical prote
32	20	100.0	149	2 T51471	farnesylated prote
33	20	100.0	152	2 T04479	cinnamyl alcohol d
34	20	100.0	152	2 H96734	probable isoprenyl
35	20	100.0	152	2 T48054	hypothetical prote
36	20	100.0	153	2 B84697	probable small hea
37	20	100.0	153	2 PN0564	von Willebrand fac
38	20	100.0	153	2 D90594	hypothetical prote
39	20	100.0	153	2 T05686	hypothetical prote
40	20	100.0	154	2 D71455	farnesylated prote
41	20	100.0	155	2 G96622	hypothetical prote
42	20	100.0	155	2 F75040	probable Heat shoc
43	20	100.0	159	2 AC2533	hypothetical prote
44	20	100.0	168	2 S73644	hypothetical prote
45	20	100.0	173	2 G64318	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S23094  
beta-amyloid protein precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 03-May-1996

C:Accession: S23094

R:Kojima, S.; Omori, M.,  
FEBS Lett. 304, 57-60, 1992

A:Title: Two-way cleavage of beta-amyloid protein precursor by multicatalytic protein

A:Reference number: S23094; MUID:92316198

A:Accession: S23094

A:Molecule type: protein

A:Residues: 1-33 <K03>

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas

Query Match 100.0%; Score 20; DB 2; Length 33;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
DB 3 VKMD 6

##### RESULT 2

RUME  
rubredoxin - Megasphaera elsdenii

C:Species: Megasphaera elsdenii

C:Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 24-Oct-1997

C:Accession: A00277

R:Baehmayer, H.; Yasunobu, K.T.; Peel, J.L.; Mayhew, S.  
J. Biol. Chem. 243, 1022-1030, 1968

A:Title: Non-heme iron proteins. V. The amino acid sequence of rubredoxin from Peptos

A:Reference number: A00277; MUID:68161650

A:Accession: A00277

A:Molecule type: protein

A:Residues: 1-52 <BAC>

C:Superfamily: rubredoxin; rubredoxin homology

C:Keywords: electron transfer; iron; metalloprotein

F:3-48/Domain: rubredoxin homology <RUB>

F:6,9,38,41/Binding site: iron (Cys) #status predicted

Query Match 100.0%; Score 20; DB 1; Length 52;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
DB 49 VKMD 52

##### RESULT 3

```

E60045
Alzheimer's disease amyloid beta/A4 protein precursor - sheep (fragment)
C:Species: Ovis sp. (sheep)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: E60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
A:Reference number: A60045; MUID:92017079
A:Accession: E60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56130
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase 1
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 57;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
    |||
    |||
Db 3 VKMD 6

RESULT 4
E60045
Alzheimer's disease amyloid beta/A4 protein precursor - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999
C:Accession: F60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
A:Reference number: A60045; MUID:92017079
A:Accession: F60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56127; NID:g1895; PIDN:CAA39592.1; PID:g1896
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase 1
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 57;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
    |||
    |||
Db 3 VKMD 6

RESULT 5
G60045
Alzheimer's disease amyloid beta/A4 protein precursor - guinea pig (fragment)
C:Species: Cavia porcellus (guinea pig)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: G60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in dog,
A:Reference number: A60045; MUID:92017079
A:Accession: G60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56126
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase 1
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 57;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4

```

```

Db 3 VKMD 6
    |||
    |||

RESULT 6
D60045
Alzheimer's disease amyloid beta/A4 protein precursor - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: D60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: D60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56124
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 57;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
    |||
    |||
Db 3 VKMD 6

RESULT 7
A60045
Alzheimer's disease amyloid beta/A4 protein precursor - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 28-Jul-1995
C:Accession: A60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: A60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56125
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 57;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
    |||
    |||
Db 3 VKMD 6

RESULT 8
B60045
Alzheimer's disease amyloid beta/A4 protein precursor - polar bear (fragment)
C:Species: Ursus maritimus (polar bear)
C:Date: 01-Dec-1992 #sequence_revision 01-Dec-1992 #text_change 13-Aug-1999
C:Accession: B60045
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.
Brain Res. Mol. Brain Res. 10, 299-305, 1991
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d
A:Reference number: A60045; MUID:92017079
A:Accession: B60045
A:Molecule type: mRNA
A:Residues: 1-57 <JOH>
A:Cross-references: EMBL:X56128; NID:g2165; PIDN:CAA39593.1; PID:g2166
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; brain

```

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 57;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
| | | |  
Db 3 VKMD 6

## RESULT 9

F96670  
hypothetical protein FL3011.6 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: F96670  
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hutzlar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: F96670  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-74 <STOP>  
A:Cross-references: GB:AE005173; NID:95042411; PIDN:AD38250.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: FL3011.6  
A:Map position: 1

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 74;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
| | | |  
Db 13 VKMD 16

## RESULT 10

A97799  
hypothetical protein RC0793 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: A97799  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: A97799  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-75 <KOR>  
A:Cross-references: GB:AE006914; PIDN:AL03331.1; PID:G15619891; GSPDB:GN00173  
C:Genetics:  
A:Gene: RC0793

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 75;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
| | | |  
Db 12 VKMD 15

## RESULT 11

P00438

Alzheimer's disease amyloid A4 protein precursor - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 30-Sep-1993 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995  
C:Accession: P00438; C60045  
R:Davidson, J.S.; West, R.L.; Kotikalapudi, P.; Maroun, L.E.  
Biochem. Biophys. Res. Commun. 188, 905-911, 1992  
A:Title: Sequence and methylation in the beta/A4 region of the rabbit amyloid precurs  
A:Reference number: P00438; MUID:93075180  
A:Accession: P00438  
A:Molecule type: DNA  
A:Residues: 1-82 <DAV>  
A:Cross-references: GB:M83558; GB:M83657  
R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.  
Brain Res. Mol. Brain Res. 10, 299-305, 1991  
A:Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d  
A:Reference number: A60045; MUID:92017079  
A:Accession: C60045  
A:Molecule type: mRNA  
A:Residues: 12-68 <JOH>  
A:Cross-references: EMBL:X56129  
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteins  
C:Keywords: alternative splicing; Alzheimer's disease; amyloid; Down's syndrome

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 82;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
| | | |  
Db 14 VKMD 17

## RESULT 12

T27174  
hypothetical protein Y54G11A.11 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27174  
R:Wallis, J.  
submitted to the EMBL Data Library, December 1998  
A:Reference number: Z20322  
A:Accession: T27174  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-84 <MLD>  
A:Cross-references: EMBL:AL034488; NID:el359895; PIDN:CAA22454.1; CESP:Y54G11A.11  
A:Experimental source: clone Y54G11A  
C:Genetics:  
A:Gene: CESP:Y54G11A.11  
A:Introns: 39/2

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 84;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
| | | |  
Db 37 VKMD 40

## RESULT 13

D98327  
cell division topological specificity factor [imported] - Agrobacterium tumefaciens (C:Species: Agrobacterium tumefaciens  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: D98327  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Iappas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A:Reference number: A97359; PMID:11743194  
A:Accession: D98327  
A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-86 <KUR>  
 A:Cross-references: GB:AE007870; PIDN:AAK90142.1; PID:q15160139; GSPDB:GN00170  
 C:Genetics:  
 A:Gene: AGR\_L\_3134  
 A:Map position: linear chromosome

Query Match 100.0%; Score 20; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 Db 60 VKMD 63

# RESULT 14

AI2955  
 cell division topological specificity factor mine [imported] - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AI2955

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
 : Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AI2955

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <KUR>

A:Cross-references: GB:AE00689, PIDN:AAL44063.1; PID:q17741627; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: mine

A:Map position: linear chromosome

Query Match 100.0%; Score 20; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 Db 60 VKMD 63

# RESULT 15

AD3625  
 cell division inhibitor mine [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AD3625

R:Delvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 .; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AD3625

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-90 <KUR>

A:Cross-references: GB:AE008918, PIDN:AAL54167.1; PID:q17985133; GSPDB:GN00191

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI10925

A:Map position: II

Query Match 100.0%; Score 20; DB 2; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 Db 63 VKMD 66

Search completed: October 30, 2002, 12:31:38  
 Job time : 1.76658 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:10 ; Search time 0.34398 Seconds

(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-103

Perfect score: 20

Sequence: 1 VKMD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	52	1 RUBR_MEGEL	P00271 megasphaera
2	20	100.0	57	1 A4_PIG	Q28023 sus scrofa
3	20	100.0	57	1 A4_URSA	Q28149 ursus marit
4	20	100.0	58	1 A4_CANFA	Q28280 canis famli
5	20	100.0	58	1 A4_RABIT	Q28748 oryctolagus
6	20	100.0	58	1 A4_SHEEP	Q28757 ovis aries
7	20	100.0	59	1 A4_BOVIN	Q28053 bos taurus
8	20	100.0	94	1 CH10_LACIA	P37283 lactococcus
9	20	100.0	103	1 TH10_LISMO	Q28386 listeria mo
10	20	100.0	134	1 VF60_ARCFU	Q28712 archaeoglob
11	20	100.0	140	1 VF62_ARCFU	Q28710 archaeoglob
12	20	100.0	168	1 VF24_MYCPN	P75254 mycoplasma
13	20	100.0	173	1 V150_METJA	Q57614 methanococ
14	20	100.0	177	1 VP22_ASFB7	P23169 african swi
15	20	100.0	179	1 VAF4_CAEEL	P52882 caenorhabdi
16	20	100.0	194	1 RS7_FUGRU	P50894 fugu rubrip
17	20	100.0	196	1 VF01_MYCPN	P52866 mycoplasma
18	20	100.0	199	1 PHB1_HUMAN	P31941 homo sapien
19	20	100.0	209	1 HOXY_ALCEU	P22319 alcalligenes
20	20	100.0	224	1 COX2_ALSCO	P48889 albinaria c
21	20	100.0	224	1 COX2_ALBTU	Q09334 albinaria t
22	20	100.0	226	1 GSHP_BOVIN	P23141 bos taurus
23	20	100.0	226	1 GSHP_HUMAN	P23142 homo sapien
24	20	100.0	226	1 GSHP_MOUSE	P46412 mus musculu
25	20	100.0	226	1 GSHP_MOUSE	P23142 mus musculu
26	20	100.0	229	1 GSHP_MOUSE	P23142 mus musculu
27	20	100.0	230	1 COX2_PARLI	P12701 paracentrot
28	20	100.0	230	1 COX2_CARAU	O78682 carassius a
29	20	100.0	230	1 COX2_CROLA	P34189 crossostoma
30	20	100.0	230	1 COX2_CYPCA	P24987 cyprinus ca
31	20	100.0	230	1 COX2_ONCMY	P48171 oncorhynchu
32	20	100.0	230	1 COX2_SALSA	O79404 salmo salar
33	20	100.0	230	1 COX2_SCYCA	O79404 scyllorhinu
			230	1 COX2_SQUAC	Q92251 squallus aca

34	20	100.0	230	1 VNS1_IACKG	P30909 influenza a
35	20	100.0	230	1 VNS1_IADA2	P03501 influenza a
36	20	100.0	230	1 VNS1_IAMA6	P13137 influenza a
37	20	100.0	230	1 VNS1_IAP11	P13141 influenza a
38	20	100.0	230	1 VNS1_IAP13	P13143 influenza a
39	20	100.0	230	1 VNS1_IATB3	P30910 influenza a
40	20	100.0	230	1 VNS1_IATKB	P30911 influenza a
41	20	100.0	230	1 VNS1_IATKC	P30912 influenza a
42	20	100.0	231	1 VNS1_IATRT	P30913 influenza a
43	20	100.0	231	1 COX2_LATCH	O03848 lactimeria c
44	20	100.0	234	1 VAI8_YEAST	P39548 saccharomyc
45	20	100.0	243	1 TONB_ENTAE	P58429 agrobacteri
					P46383 enterobacte

## ALIGNMENTS

RESULT 1	
RUBR_MEGEL	
ID P00271.	STANDARD: PRT; 52 AA.
AC	21-JUL-1986 (Rel. 01, Created)
DT	21-JUL-1986 (Rel. 01, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Rubredoxin (rd).
OS	Megasphaera elsdenii.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Sporomusa subbranch;
OC	Megasphaera.
OX	NCBI_TaxID=907;
RN	[1]
RP	SEQUENCE.
RX	MEDLINE=68161650; PubMed=5640967;
RA	Bachmayer H., Yasunobu K.T., Peel J.L., Mayhew S.G.;
RT	"Non-heme iron proteins. V. The amino acid sequence of rubredoxin
RT	from peptostreptococcus elsdenii.";
RL	J. Biol. Chem. 243:1022-1030(1968).
CC	-1- FUNCTION: RUBREDOXIN IS A SMALL, NONHEME, IRON PROTEIN LACKING
CC	ACID-LABILE SULFIDE. ITS SINGLE FE, CHELATED TO 4CYS, FUNCTIONS
CC	AS AN ELECTRON ACCEPTOR AND MAY ALSO STABILIZE THE CONFORMATION
CC	OF THE MOLECULE.
CC	-1- SIMILARITY: BELONGS TO THE RUBREDOXIN FAMILY.
CC	PIR: A00277; RUME.
DR	HSP: P00269; IRB9.
DR	InterPro: IPR004039; Rubredox.
DR	InterPro: IPR001052; Rubredoxin.
DR	Pfam: PF00301; rubredoxin.1.
DR	PRINTS: PR00163; RUBREDOXIN.
DR	ProDom: PD001610; Rubredoxn_domain; 1.
DR	PROSITE: PS00202; RUBREDOXIN.1.
KW	Electron transport; Iron; Metal-binding.
FT	METAL 6 6 IRON.
FT	METAL 9 9 IRON.
FT	METAL 38 38 IRON.
FT	METAL 41 41 IRON.
SQ	SEQUENCE 52 AA; 5616 MW; A13AB077992PDPD4 CRC64;
QY	1 VKMD 4
DB	49 VKMD 52
QY	1 VKMD 4
DB	49 VKMD 52
RESULT 2	
A4_PIG	STANDARD: PRT; 57 AA.
AC	Q29023;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid

```

DE protein (Beta-APP) (A-beta)] (Fragment).
GN APP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: X56127; CAA39592.1; -.
CC HSSP: P05067; IBA4.
CC InterPro: IPR001868; A4_APP. PARTIAL.
CC PROSITE: PS00319; A4_EXTRA; PARTIAL.
CC PROSITE: PS00320; A4_INTRA; PARTIAL.
CC Glycoprotein; Amyloid; Neurone; Transmembrane.
CC NON_TER 1 1
CC CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
CC DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 34 57 POTENTIAL.
CC NON_TER 57 57
CC SEQUENCE 57 AA; 6172 MW; 84209D88EBA82DFA CRC64;

Query Match 100.0%; Score 20; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
   ||||
Db 3 VKMD 6

RESULT 3
A4_URSMA STANDARD; PRT; 57 AA.
AC 029149;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alzheimer's disease amyloid A4 protein homolog [contains: Beta-amyloid
DE protein (Beta-APP) (A-beta)] (Fragment).
GN APP.
OS Ursus maritimus (Polar bear) (Thalartos maritimus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Ursidae; Ursus.
OX NCBI_TaxID=29073;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";

```

```

RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: X56128; CAA39593.1; -.
CC HSSP: P05067; IAML.
CC InterPro: IPR001868; A4_APP.
CC PROSITE: PS00319; A4_EXTRA; PARTIAL.
CC PROSITE: PS00320; A4_INTRA; PARTIAL.
CC Glycoprotein; Amyloid; Neurone; Transmembrane.
CC NON_TER 1 1
CC CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
CC DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 34 57 POTENTIAL.
CC NON_TER 57 57
CC SEQUENCE 57 AA; 6172 MW; 84209D88EBA82DFA CRC64;

Query Match 100.0%; Score 20; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
   ||||
Db 3 VKMD 6

RESULT 4
A4_CANFA STANDARD; PRT; 58 AA.
AC 026280;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alzheimer's disease amyloid A4 protein homolog [contains: Beta-amyloid
DE protein (Beta-APP) (A-beta)] (Fragment).
GN APP.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```
CC -----
DR EMBL: X56125; CAA39590.1; -
DR HSSP: P05067; 1BA4.
DR InterPro: IPR001868; A4_APP.
DR PROSITE: PS00319; A4_EXTRA; PARTIAL.
DR PROSITE: PS00320; A4_INTRA; PARTIAL.
DR Glycoprotein: Amyloid; Neurone; Transmembrane.
KW NON_TER 1 1
FT CHAIN 7 49 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 34 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 35 58 POTENTIAL.
FT NON_TER 58 58
SQ SEQUENCE 58 AA: 6285 MW: 8469D488A2E12DFA CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
   |||
Db 4 VKMD 7

RESULT 5
A4_RABIT STANDARD: PRT; 58 AA.
ID A4_RABIT
AC Q28748;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid
DE protein (beta-APP) (A-beta)] (Fragment).
GN App.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56129; CAA39594.1; -
DR HSSP: P05067; 1BA4.
DR InterPro: IPR001868; A4_APP.
DR PROSITE: PS00319; A4_EXTRA; PARTIAL.
DR PROSITE: PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein: Amyloid; Neurone; Transmembrane.
KW NON_TER 1 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
SQ SEQUENCE 58 AA: 6300 MW: F434209D88BBA82D CRC64;
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
   |||
Db 3 VKMD 6

RESULT 6
A4_SHEEP STANDARD: PRT; 58 AA.
ID A4_SHEEP
AC Q28757;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alzheimer's disease amyloid A4 protein homolog [Contains: Beta-amyloid
DE protein (beta-APP) (A-beta)] (Fragment).
GN App.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_Taxid=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=92017079; PubMed=1656157;
RA Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT "Conservation of the sequence of the Alzheimer's disease amyloid
RT peptide in dog, polar bear and five other mammals by cross-species
RT polymerase chain reaction analysis.";
RL Brain Res. Mol. Brain Res. 10:299-305(1991).
CC -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56130; CAA39595.1; -
DR HSSP: P05067; 1AHL.
DR InterPro: IPR001868; A4_APP.
DR PROSITE: PS00319; A4_EXTRA; PARTIAL.
DR PROSITE: PS00320; A4_INTRA; PARTIAL.
KW Glycoprotein: Amyloid; Neurone; Transmembrane.
KW NON_TER 1 1
FT CHAIN 6 48 BETA-AMYLOID PROTEIN (POTENTIAL).
FT DOMAIN <1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 57 POTENTIAL.
FT DOMAIN 58 >58 CYTOPLASMIC (POTENTIAL).
FT NON_TER 58 58
SQ SEQUENCE 58 AA: 6300 MW: F434209D88BBA82D CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4
   |||
Db 3 VKMD 6

RESULT 7
A4_BOVIN
```

```

ID      A4_BOVIN          STANDARD:          PRT:          59 AA.
AC      Q28053;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Alzheimer's disease amyloid A4 protein homolog [contains: Beta-amyloid
DE      protein (Beta-ApP) (A-beta)] (Fragment).
OS      APP.
GN      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=92017079; PubMed=1656157;
RT      Johnstone E.M., Chaney M.O., Norris F.H., Pascual R., Little S.P.;
RT      "Conservation of the sequence of the Alzheimer's disease amyloid
RT      peptide in dog, polar bear and five other mammals by cross-species
RT      polymerase chain reaction analysis.";
RL      Brain Res. Mol. Brain Res. 10:299-305(1991).
CC      -1- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC      INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC      G1O (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X56124; CA39589.1; -
DR      EMBL: X56126; CA39591.1; -
DR      HSSP: P05067; 1BA4.
DR      InterPro: IPR001868; A4_APP.
DR      PROSITE: PS00319; A4_EXTRA; PARTIAL.
DR      PROSITE: PS00320; A4_INTRA; PARTIAL.
KW      Glycoprotein; Amyloid; Neurone; Transmembrane.
FT      NON_TER          1          1
FT      CHAIN            1          49      BETA-AMYLOID PROTEIN (POTENTIAL).
FT      DOMAIN          <1          34      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        35          58      POTENTIAL.
FT      DOMAIN          59          59      CYTOPLASMIC (POTENTIAL).
FT      NON_TER          59          59
SQ      SEQUENCE      59 AA: 6414 MW: F43469D48A2E12D CRC64;

Query Match          100.0%; Score 20; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 34;
Matches      4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 VKMD 4
        ||||
DB      4 VKMD 7

RESULT 8
CH10_LACLA          STANDARD:          PRT:          94 AA.
ID      CH10_LACLA
AC      P37283;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      10-kDa chaperonin (Protein Cpn10) (groES protein).
GN      GROES OR GROES OR IL0393.
OS      Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC      Lactococcus.
OX      NCBI_TaxID=1360;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93252268; PubMed=8486277;
RA      Kim S.G., Batt C.A.;
RT      "Cloning and sequencing of the Lactococcus lactis subsp. lactis
RT      groESL operon.";
RL      Gene 127:121-126(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=IL1403;
RX      MEDLINE=21235186; PubMed=11337471;
RA      Bolotin A., Wincker P., Manger S., Jallou O., Malame K.,
RA      Weissenbach J., Ehrlich S.D., Sorokin A.;
RT      "The complete genome sequence of the lactic acid bacterium Lactococcus
RT      lactis ssp. lactis IL1403.";
RL      Genome Res. 11:731-753(2001).
CC      -1- FUNCTION: BINDS TO CPN60 IN THE PRESENCE OF MG-ATP AND SUPPRESSES
CC      THE ATPASE ACTIVITY OF THE LATTER.
CC      -1- SUBUNIT: HEPTAMER OF 7 SUBUNITS ARRANGED IN A RING
CC      (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE GROES CHAPERONIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X71132; CA50445.1; -
DR      EMBL: AE006276; AAK0491.1; -
DR      PIR: S32105; S32105.
DR      PIR: JN0660; JN0660.
DR      InterPro: IPR001476; Cpn10.
DR      Pfam: PF00166; Cpn10; 1.
DR      PRINTS: PR00297; CHAPERONIN10.
DR      PRODOM: PD00566; Cpn10; 1.
DR      PROSITE: PS00681; CHAPERONINS_CPN10; 1.
KW      Chaperone; Complete proteome.
SQ      SEQUENCE      94 AA: 10221 MW: 291913B147CA66B CRC64;

Query Match          100.0%; Score 20; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 54;
Matches      4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 VKMD 4
        ||||
DB      74 VKMD 77

RESULT 9
TH10_LISMO          STANDARD:          PRT:          103 AA.
ID      TH10_LISMO
AC      Q95386;
DT      01-MAR-2002 (Rel. 41, Created)
DT      01-MAR-2002 (Rel. 41, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Thioedoxin (Trx).
GN      TRXA OR LMO1233 OR LIN1196.
OS      Listeria monocytogenes, and
OS      Listeria innocua.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Listeria.
OX      NCBI_TaxID=1639, 1642;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SPECIES=L.monocytogenes; STRAIN=BGD;
RA      Tarchanov M., Borovok I., Aharonowitz Y., Cohen G.;
RT      "Isolation, cloning and characterization of the Listeria monocytogenes
RT      thioedoxin gene, trxA.";
RL      Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
RN      [2]

```



RA SEQUENCE FROM N.A.  
 RC SPECIES=L. monocytogenes, and L. innocua;  
 RX STRAIN=ED-e / Serovar 1/2a, and CLIP 11262 / Serovar 6a;  
 RX MEDLINE=11537279; PubMed=11679669;  
 RA Glaeser P., Frangoul L., Buchrieser C., Rusnlok C., Amend A.,  
 RA Baquero F., Berche P., Bioecker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Deloux P.,  
 RA Damm E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,  
 RA Entian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hahn T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madano E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstedt G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schueter T., Simoes N., Tietze A.,  
 RA Vazquez-Boland J.-A., Voss H., Wenland J., Cossart P.;  
 RX Science 294:849-852(2001).  
 CC "Comparative genomics of *Listeria* species."  
 CC -1- FUNCTION: COMPONENT OF THE THIOREDOXIN-THIOREDOXIN REDUCTASE  
 CC SYSTEM. THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS  
 CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL, TO  
 CC A DISULFIDE, AND CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ133066; CAB40815.2; -  
 DR EMBL: AL591978; CAC93311.1; -  
 DR EMBL: AL596167; CAC96427.1; -  
 DR HSSP: P80579; 10UW.  
 DR Listlist: LM01196; -  
 DR Listlist: LM011233; -  
 DR InterPro: IPR000063; ThioRed.  
 DR Pfam: PF00085; thioRed; 1.  
 DR PRINTS: PR00421; THIOREDOXIN.  
 DR PROSITE: PS00194; THIOREDOXIN; 1.  
 DR Oxidoreductase: Redox-active center: Complete proteome.  
 KW DISULFID 28 31 REDOX-ACTIVE (BY SIMILARITY).  
 FT SEQUENCE 103 AA: 11620 MW; 01F6A77434559A64 CRC64;  
 SQ  
 Query Match 100.0%; Score 20; DB 1; Length 103;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VKMD 4  
 Db 52 VKMD 55  
 RESULT 10  
 YF62\_ARCFU STANDARD; PRT; 134 AA.  
 ID YF62\_ARCFU  
 AC 028712;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AFI560.  
 GN AFI560.  
 OS Archaeoglobus fulgidus.  
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus.  
 CC NCB1\_TaxID=2234;  
 RN [1]  
 RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriakos N.C.,  
 RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujil C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RX Science 294:364-370(1997).  
 CC "The complete genome sequence of the hyperthermophilic, sulphate-  
 CC reducing archaeon *Archaeoglobus fulgidus*."  
 CC RT Nature 390:364-370(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000995; AAB89698.1; -  
 DR TIGR: AF1560; -  
 KW Hypothetical protein: Transmembrane: Complete proteome.  
 FT TRANSMEM 13 35 POTENTIAL.  
 SQ SEQUENCE 134 AA: 15036 MW; EEB83D22A904A79 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VKMD 4  
 Db 83 VKMD 86  
 RESULT 11  
 YF62\_ARCFU STANDARD; PRT; 140 AA.  
 ID YF62\_ARCFU  
 AC 028710;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AFI562.  
 GN AFI562.  
 OS Archaeoglobus fulgidus.  
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 CC Archaeoglobus.  
 CC NCB1\_TaxID=2234;  
 RN [1]  
 RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriakos N.C.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujil C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RX Science 294:364-370(1997).  
 CC "The complete genome sequence of the hyperthermophilic, sulphate-  
 CC reducing archaeon *Archaeoglobus fulgidus*."  
 CC RT Nature 390:364-370(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000994; AAB89687.1; -
CC DR TIGR; AF1562; -
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC FT TRANSMEM 20 42 POTENTIAL.
CC FT TRANSMEM 88 110 POTENTIAL.
CC FT TRANSMEM 115 137 POTENTIAL.
CC SO SEQUENCE 140 AA; 15667 MW; 937DCB585A17991 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 140;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
   ||||
Db 2 VKMD 5

RESULT 12
VF24_MYCPN STANDARD; PRT; 168 AA.
ID VF24_MYCPN
AC P75254;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN524 (G12_Orf168).
GN MPN524 OR MP318.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=2104;
   [1]
   SEQUENCE FROM N.A.
   RC STRAIN=ATCC 29342 / M129;
   RX MEDLINE=97105885; PubMed=8948633;
   RA Himmelfeich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
   RA Herrmann R.;
   RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
   RT pneumoniae.";
   RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE UPF0134 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000029; AAB95966.1; -
CC DR InterPro; IPR002862; DUF16.
CC PFam; PF01519; DUF16; 1.
CC KW Hypothetical protein; Complete proteome.
CC SO SEQUENCE 168 AA; 20110 MW; 0878B81DEB019170 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 168;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
   ||||
Db 125 VKMD 128

RESULT 13
VF150_METJA STANDARD; PRT; 173 AA.
ID VF150_METJA
AC Q57614;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0150.
GN MJ0150.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
CC NCBI_TaxID=2190;
   [1]
   SEQUENCE FROM N.A.
   RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
   RX MEDLINE=96337999; PubMed=8688087;
   RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
   RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
   RA Kerevage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
   RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
   RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fuhrmann J.L., Nguyen D.,
   RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
   RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
   RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
   RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
   RT jannaschii".
   RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE M.JANNASCHII MJ0150 / MJ0739 / MJ0745 /
CC MJ1460 / MJ1642 FAMILY AND TO THE C-TERMINAL OF M.JANNASCHII
CC MJ0855.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67472; AAB98141.1; -
CC DR TIGR; MJ0150; -
CC DR InterPro; IPR004096; VAR.
CC PFam; PF02830; VAR; 1.
CC KW Hypothetical protein; Complete proteome.
CC SO SEQUENCE 173 AA; 19524 MW; A273F382C91D0315 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 1; Length 173;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
   ||||
Db 4 VKMD 7

RESULT 14
VF22_ASFB7 STANDARD; PRT; 177 AA.
ID VF22_ASFB7
AC P23169;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Protein P22 precursor (protein K177).
DE African swine fever virus (strain BA71V) (ASFV).
OS African swine fever virus; no RNA stage; Asfarviridae;
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
CC African swine fever-like viruses.
CC NCBI_TaxID=10458;
   [1]
   SEQUENCE FROM N.A.
   RP MEDLINE=90219205; PubMed=2325203;
   RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
   RA de la Vega I., Blasco R., Vinuela E.;
   RT "Multigene families in African swine fever virus: family 360.";
   RL J. Virol. 64:2073-2081(1990).
   [2]
   CHARACTERIZATION.
   RP

```

RX MEDLINE-91134988: PubMed-1994575;

RA Camacho A., Vinuela E.;

RT "Protein p22 of African swine fever virus: an early structural

RL protein that is incorporated into the membrane of infected cells.";

CC Virology 181:251-257(1991).

CC -I- FUNCTION: EARLY STRUCTURAL PROTEIN THAT IS INCORPORATED INTO THE

CC MEMBRANE OF INFECTED CELLS.

CC -I- SUBCELLULAR LOCATION: PROBABLE EXTERNAL VIRAL STRUCTURAL PROTEIN

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U18466; AAA65240.1; -

DR EMBL: M57546; AAA42681.1; -

KW Early protein; Signal; Membrane; Structural protein.

FT SIGNAL 1 27

FT CHAIN 28 177

FT POTENTIAL. P22.

SO SEQUENCE 177 AA; 20160 MW; 7C2EE66F13E762D1 CRC64;

Query Match

Best Local Similarity 100.0%; Score 20; DB 1; Length 177;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4

Db 65 VKMD 68

RESULT 15

YAF4\_CAEEL

ID YAF4\_CAEEL

AC P52882;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Hypothetical 20.4 kDa protein F46C5.4 in chromosome II.

GN F46C5.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Thomas K.;

RU Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: 254281; CA91046.1; -

DR Wormpep: F46C5.4; CE03346.

KW Hypothetical protein.

SO SEQUENCE 179 AA; 20375 MW; 4662982743926408 CRC64;

Query Match

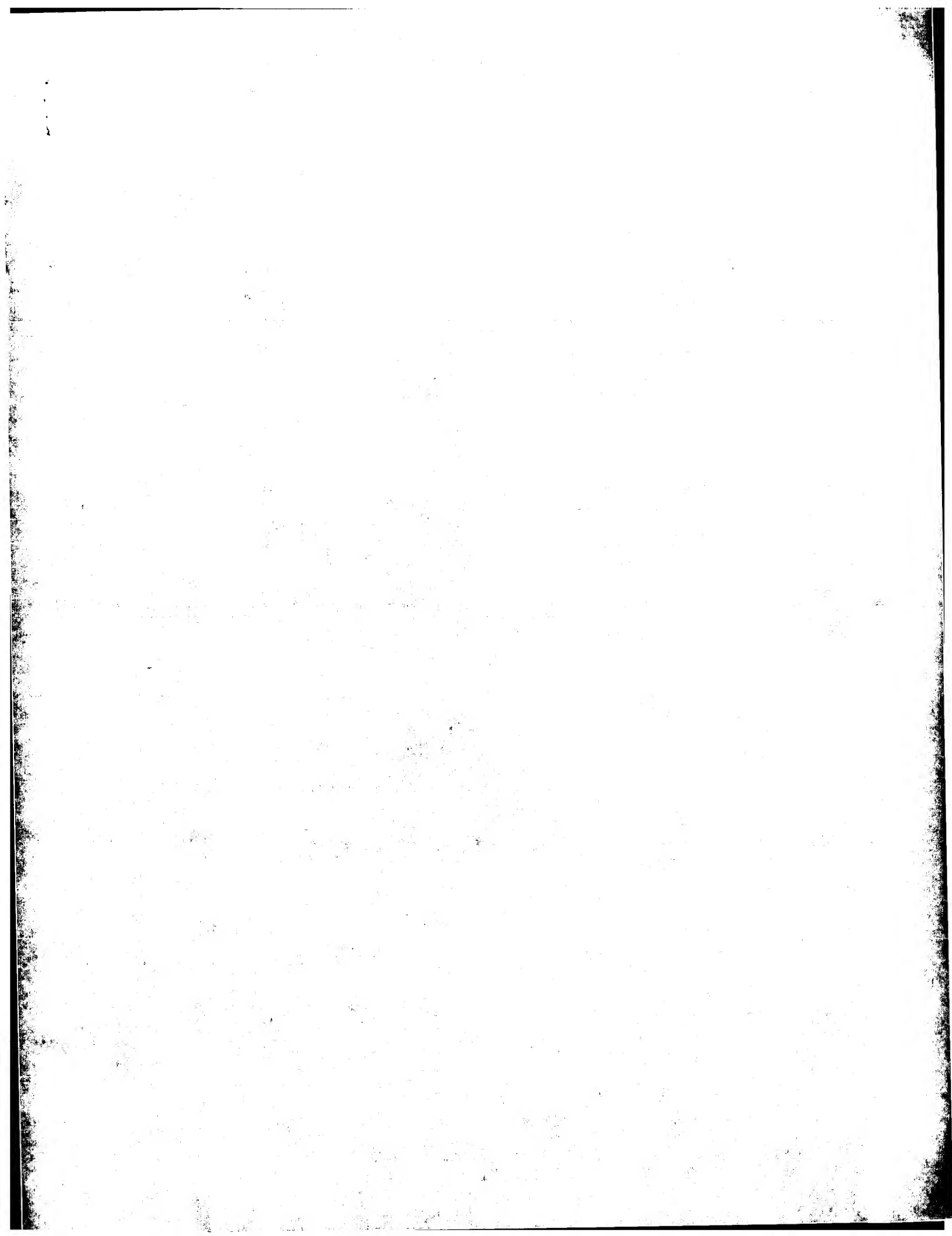
Best Local Similarity 100.0%; Score 20; DB 1; Length 179;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4

Db 119 VKMD 122

Search completed: October 30, 2002, 12:27:56  
Job time : 2.34398 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 1.24816 Seconds  
(without alignments)  
554.401 Million cell updates/sec

Title: US-09-724-571-103

Perfect score: 20

Sequence: 1 VKMD 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	% Match	Query Length	DB ID	Description
1	20	100.0	60	5 Q9BMS3	Q9BMS3 plasmodium
2	20	100.0	62	5 Q9BJU6	Q9BJU6 plasmodium
3	20	100.0	62	5 Q9BJU4	Q9BJU4 plasmodium
4	20	100.0	65	11 054696	054696 mus musculus
5	20	100.0	74	10 Q9XIR8	Q9XIR8 arabidopsis
6	20	100.0	75	16 Q92HH8	Q92HH8 rickettsia
7	20	100.0	79	11 035463	035463 cricetus
8	20	100.0	82	4 P78438	P78438 homo sapien
9	20	100.0	82	4 Q9U177	Q9U177 homo sapien
10	20	100.0	82	4 Q16014	Q16014 homo sapien
11	20	100.0	82	4 Q16019	Q16019 homo sapien
12	20	100.0	82	4 Q16020	Q16020 homo sapien
13	20	100.0	82	5 Q9U501	Q9U501 manduca sex
14	20	100.0	83	4 Q96I14	Q96I14 homo sapien
15	20	100.0	84	5 Q9XVZ8	Q9XVZ8 ctenorhachi
16	20	100.0	86	1 Q9C4V4	Q9C4V4 sulfolobus

17	20	100.0	90	11 P70443	P70443 mus musculus
18	20	100.0	92	9 Q64108	Q64108 bacterioph
19	20	100.0	92	16 Q34730	Q34730 bacillus su
20	20	100.0	93	5 Q21047	Q21047 caenorhachi
21	20	100.0	94	2 Q9AEP8	Q9AEP8 lactococcus
22	20	100.0	98	5 Q9GPA5	Q9GPA5 drosophila
23	20	100.0	101	12 Q91ZP8	Q91ZP8 tick-borne
24	20	100.0	103	16 Q9S386	Q9S386 listeria mo
25	20	100.0	104	5 Q9GP75	Q9GP75 drosophila
26	20	100.0	104	5 Q9GP72	Q9GP72 drosophila
27	20	100.0	104	5 Q9GN84	Q9GN84 drosophila
28	20	100.0	104	5 Q9GN83	Q9GN83 drosophila
29	20	100.0	104	5 Q9GN82	Q9GN82 drosophila
30	20	100.0	104	16 Q97PE8	Q97PE8 streptococc
31	20	100.0	105	5 Q9GP79	Q9GP79 drosophila
32	20	100.0	105	5 Q9GP78	Q9GP78 drosophila
33	20	100.0	105	5 Q9GNH2	Q9GNH2 drosophila
34	20	100.0	106	5 Q9GP80	Q9GP80 drosophila
35	20	100.0	115	15 Q74763	Q74763 human immun
36	20	100.0	125	12 Q910P5	Q910P5 tick-borne
37	20	100.0	127	17 Q97X15	Q97X15 sulfolobus
38	20	100.0	137	12 Q83000	Q83000 bunyavirus
39	20	100.0	137	12 Q91ZP1	Q91ZP1 tick-borne
40	20	100.0	137	12 Q91E36	Q91E36 rhocid herp
41	20	100.0	138	10 Q9ZRP7	Q9ZRP7 glycine max
42	20	100.0	142	12 Q91Z44	Q91Z44 tick-borne
43	20	100.0	142	16 Q9K7K4	Q9K7K4 bacillus ha
44	20	100.0	145	12 Q91MX8	Q91MX8 lumpy skin
45	20	100.0	145	17 Q9V2N3	Q9V2N3 pyrococcus

## ALIGNMENTS

RESULT 1  
ID Q9BMS3 PRELIMINARY: PRT: 60 AA.  
AC Q9BMS3;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE ERYTHROCYTE MEMBRANE PROTEIN 1 (FRAGMENT).  
GN VARL.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=56933;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21273141; PubMed=11378202;  
RA Noviyanti R., Brown G.V., Wickham M.E., Duffy M.F., Cowman A.F.,  
RA Reeder J.C.;  
RT "Multiple var gene transcripts are expressed in Plasmodium falciparum  
RT infected erythrocytes selected for adhesion.";  
RL Mol. Biochem. Parasitol. 114:227-237(2001).  
DR EMBL; AF306406; AAK00867.1; -;  
FT NON\_TER 1  
FT SEQUENCE 60 AA; 6798 MW; 83776F68E9DB9291 CRC64;

Query Match 100.0%; Score 20; DB 5; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VKMD 4  
Db 6 VKMD 9

RESULT 2  
ID Q9BJU6 PRELIMINARY: PRT: 62 AA.  
AC Q9BJU6;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)

```

DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
DE VARIANT SURFACE PROTEIN (FRAGMENT).
CN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=93;
RA Nogueira P.A., Wunderlich G., Tada M.S., Costa J.D.N., Menezes M.J.,
RA Scherf A., Pereira da Silva L.H.;
RT "Plasmodium falciparum: repertoire of expressed var genes and adhesion
RT properties to endothelial receptors of clinical isolates from patients
RT in Rondonia (Brazilian western Amazon region).";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF333282; AAK19553.1; -.
DR InterPro: IPR001230; Prenyltin.
DR PROSITE: PS00294; PRENYLTATION; UNKNOWN_1.
FT NON_TER 1 62
FT NON_TER 1 62
SQ SEQUENCE 62 AA; 6953 MW; 83209EA45C1D57A5 CRC64;

Query Match 100.0%; Score 20; DB 5; Length 62;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
DB 10 VKMD 13

RESULT 3
O9BJJ4 PRELIMINARY; PRT; 62 AA.
ID O9BJJ4;
AC O9BJJ4;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
DE VARIANT SURFACE PROTEIN (FRAGMENT).
CN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=95;
RA Nogueira P.A., Wunderlich G., Tada M.S., Costa J.D.N., Menezes M.J.,
RA Scherf A., Pereira da Silva L.H.;
RT "Plasmodium falciparum: repertoire of expressed var genes and adhesion
RT properties to endothelial receptors of clinical isolates from patients
RT in Rondonia (Brazilian western Amazon region).";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF333284; AAK19555.1; -.
DR InterPro: IPR001230; Prenyltin.
DR PROSITE: PS00294; PRENYLTATION; UNKNOWN_1.
FT NON_TER 1 62
FT NON_TER 1 62
SQ SEQUENCE 62 AA; 7029 MW; F55E3AEBAFBADCAC CRC64;

Query Match 100.0%; Score 20; DB 5; Length 62;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
DB 10 VKMD 13

RESULT 4
O54696 PRELIMINARY; PRT; 65 AA.
ID O54696;
AC O54696;

```

```

DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE PROSTAGLANDIN F2 ALPHA RECEPTOR REGULATORY PROTEIN (FRAGMENT).
CN PGFRN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97109715; Pubmed=8951995;
RX Orlicky D.J., Nordeen S.K.;
RA "Cloning, sequencing and proposed structure for a prostaglandin F2
RT alpha receptor regulatory protein.";
RL Prostaglandins Leukot. Essent. Fatty Acids 55:261-268(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Orlicky D.J.;
RT "Mouse prostaglandin F2 alpha receptor regulatory protein cDNA
RT fragment.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF006201; AAB87643.1; -.
DR MGD: MGI:1277114; Pgfrn.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 1 1
SQ SEQUENCE 65 AA; 7623 MW; 69D3E97775D86729 CRC64;

Query Match 100.0%; Score 20; DB 11; Length 65;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
DB 7 VKMD 10

RESULT 5
O9XIR8 PRELIMINARY; PRT; 74 AA.
ID O9XIR8;
AC O9XIR8;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE F13011.6 PROTEIN.
DE F13011.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altati H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaya V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006193; AAD38250.1; -.
SQ SEQUENCE 74 AA; 8859 MW; F76AAF00BEB8818A CRC64;

Query Match 100.0%; Score 20; DB 10; Length 74;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4
DB 13 VKMD 16

RESULT 6

```

092HH8  
ID Q92HH8 PRELIMINARY; PRT; 75 AA.  
AC Q92HH8;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DR HYPOTHETICAL PROTEIN RC0793.  
GN RC0793.  
OS Rickettsia conorii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.  
OX NCBI\_Taxid=781;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ALISH 7.  
RX MEDLINE=21442074; PubMed=11557893;  
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-F., Barbe V.,  
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
RA Raoult D.;  
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
RL Science 293:2093-2098(2001).  
DR EMBL; AE008635; AL03331.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 75 AA; 8841 MW; 66EE9C73C0D29592 CRC64;  
  
Query Match  
Best Local Similarity 100.0%; Score 20; DB 16; Length 75;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VKMD 4  
DB 12 VKMD 15  
  
RESULT 7  
ID Q35463 PRELIMINARY; PRT; 79 AA.  
AC Q35463;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DR ALZHEIMER'S AMYLOID BETA PROTEIN (FRAGMENT).  
GN BETA APP.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
OX NCBI\_Taxid=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sambamurti K., Plunik I., Gandhi S.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF030413; AAB86608.1; -;  
DR HSSP; P05067; 1BA4.  
FT NON\_TER 1 79  
FT NON\_TER 79 79  
SQ SEQUENCE 79 AA; 8538 MW; 37FC6C3BF3F597 CRC64;  
  
Query Match  
Best Local Similarity 100.0%; Score 20; DB 11; Length 79;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VKMD 4  
DB 18 VKMD 21  
  
RESULT 8  
ID P78438 PRELIMINARY; PRT; 82 AA.  
AC P78438;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE AMYLOID PROTEIN (BETA-AMYLOID PROTEIN) (FRAGMENT).  
GN APP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=89392030; PubMed=2675837;  
RA Johnstone E.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H.,  
RA Little S.P.;  
RT "Alzheimer's disease amyloid peptide is encoded by two exons and shows  
RT similarity to soybean trypsin inhibitor.";  
RL Biochem. Biophys. Res. Commun. 163:1248-1255(1989).  
RN [2]  
RP SEQUENCE OF 19-48 FROM N.A.  
RX MEDLINE=87120329; PubMed=2949367;  
RA Tanzi R.E., Gusella J.F., Watkins P.C., Bruns G.A., George-Hyslop P.,  
RA Van Keuren M.L., Patterson D., Pagan S., Kurnit D.M., Neve R.L.;  
RT "Amyloid beta protein gene: cDNA, mRNA distribution, and genetic  
RL linkage near the Alzheimer locus.";  
RN [3]  
RP SEQUENCE OF 32-63 FROM N.A.  
RX MEDLINE=93035397; PubMed=1415269;  
RA Kawano K., Orr H.T., Payaml H., Wijsman E.M., Alonso M.E., Pulst S.M.,  
RA Anderson L., O' Dahl S., Nemens E., White J.A.;  
RT "Linkage and mutational analysis of familial Alzheimer disease  
RT kinds for the APP gene region.";  
RL Am. J. Hum. Genet. 51:998-1014(1992).  
DR EMBL; M29270; AAS1768.1; -;  
DR EMBL; M29269; AAS1768.1; JOINED.  
DR EMBL; M15532; AAS1564.1; -;  
DR EMBL; S45136; AAB23646.1; -;  
DR HSSP; P05067; 1BA4.  
FT NON\_TER 1 1  
FT NON\_TER 1 1  
SQ SEQUENCE 82 AA; 8994 MW; 8DA9EA2B813A070E CRC64;  
  
Query Match  
Best Local Similarity 100.0%; Score 20; DB 4; Length 82;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VKMD 4  
DB 14 VKMD 17  
  
RESULT 9  
ID Q9U177 PRELIMINARY; PRT; 82 AA.  
AC Q9U177;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DR PRO214 PROTEIN (DJ319D22.2).  
GN DJ319D22.2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,  
RA Liu M., He F.;  
RT "Functional prediction of the coding sequences of 50 new genes deduced  
RT by analysis of cDNA clones from human fetal liver";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Williams S.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF090903: AAF24027.1; -  
 SQ EMBL: AL133262; CAC08558.1; -  
 SEQUENCE 82 AA: 9443 MW: E29D37E1E2B327A1 CRC64;

Query Match 100.0%; Score 20; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 DB 19 VKMD 22

## RESULT 10

Q16014 PRELIMINARY; PRT; 82 AA.  
 AC Q16014;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-AMYLLOID PEPTIDE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93236601; PubMed=8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease  
 mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL: S60721; AAB26263.2; -  
 DR HSSP: P05067; 1BA4.  
 FT NON\_TER 1 82  
 FT SEQUENCE 82 AA: 8972 MW: F534AA5B3EA9230A CRC64;  
 SQ

Query Match 100.0%; Score 20; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 DB 15 VKMD 18

## RESULT 11

Q16019 PRELIMINARY; PRT; 82 AA.  
 ID Q16019  
 AC Q16019;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-AMYLLOID PEPTIDE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93236601; PubMed=8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease  
 mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL: S61380; AAB26264.2; -  
 DR HSSP: P05067; 1BA4.  
 FT NON\_TER 1 82  
 FT SEQUENCE 82 AA: 8938 MW: F534AA50E579230A CRC64;  
 SQ

Query Match 100.0%; Score 20; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 DB 15 VKMD 18

## RESULT 12

Q16020 PRELIMINARY; PRT; 82 AA.  
 ID Q16020  
 AC Q16020;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-AMYLLOID PEPTIDE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93236601; PubMed=8476439;  
 RA Denman R.B., Rosenzwaig R., Miller D.L.;  
 RT "A system for studying the effect(s) of familial Alzheimer disease  
 mutations on the processing of the beta-amyloid peptide precursor.";  
 RL Biochem. Biophys. Res. Commun. 192:96-103(1993).  
 DR EMBL: S61383; AAB26263.2; -  
 DR HSSP: P05067; 1BA4.  
 FT NON\_TER 1 82  
 FT SEQUENCE 82 AA: 8882 MW: F534AA5AE5D9230A CRC64;  
 SQ

Query Match 100.0%; Score 20; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKMD 4  
 ||||  
 DB 15 VKMD 18

## RESULT 13

Q9U501 PRELIMINARY; PRT; 82 AA.  
 ID Q9U501  
 AC Q9U501;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 9.5 KDA PROTEIN.  
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Lepidoptera; Glossata; Ditrysia;  
 OC Sphingioidea; Sphingidae; Sphinginae; Manduca.  
 NX NCBI\_TaxID=7130;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20099029; PubMed=10620045;  
 RA Robertson H.M., Martos R., Sears C., Todres E.Z., Walden K.O.,  
 RA Nard J.B.;  
 RT "Diversity of odourant binding proteins revealed by an expressed  
 sequence tag project on male Manduca sexta moth antennae.";  
 RL Insect Mol. Biol. 8:501-518(1999).  
 DR EMBL: AF117587; AAF16709.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 82 AA: 9460 MW: AB889CE023AB4E3C CRC64;

Query Match 100.0%; Score 20; DB 5; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY 1 VKMD 4  
 ||||  
 Db 37 VKMD 40

## RESULT 14

096114 PRELIMINARY; PRT; 83 AA.  
 AC 096114;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DE SIMILAR TO CG6244 GENE PRODUCT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, AND RHADOMYOSARCOMA;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BC007516; AA07516.1; -  
 SO SEQUENCE 83 AA; 9462 MW; 77F1EBEC72AC6F CRC64;

Query Match 100.0%; Score 20; DB 4; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
 ||||  
 Db 37 VKMD 40

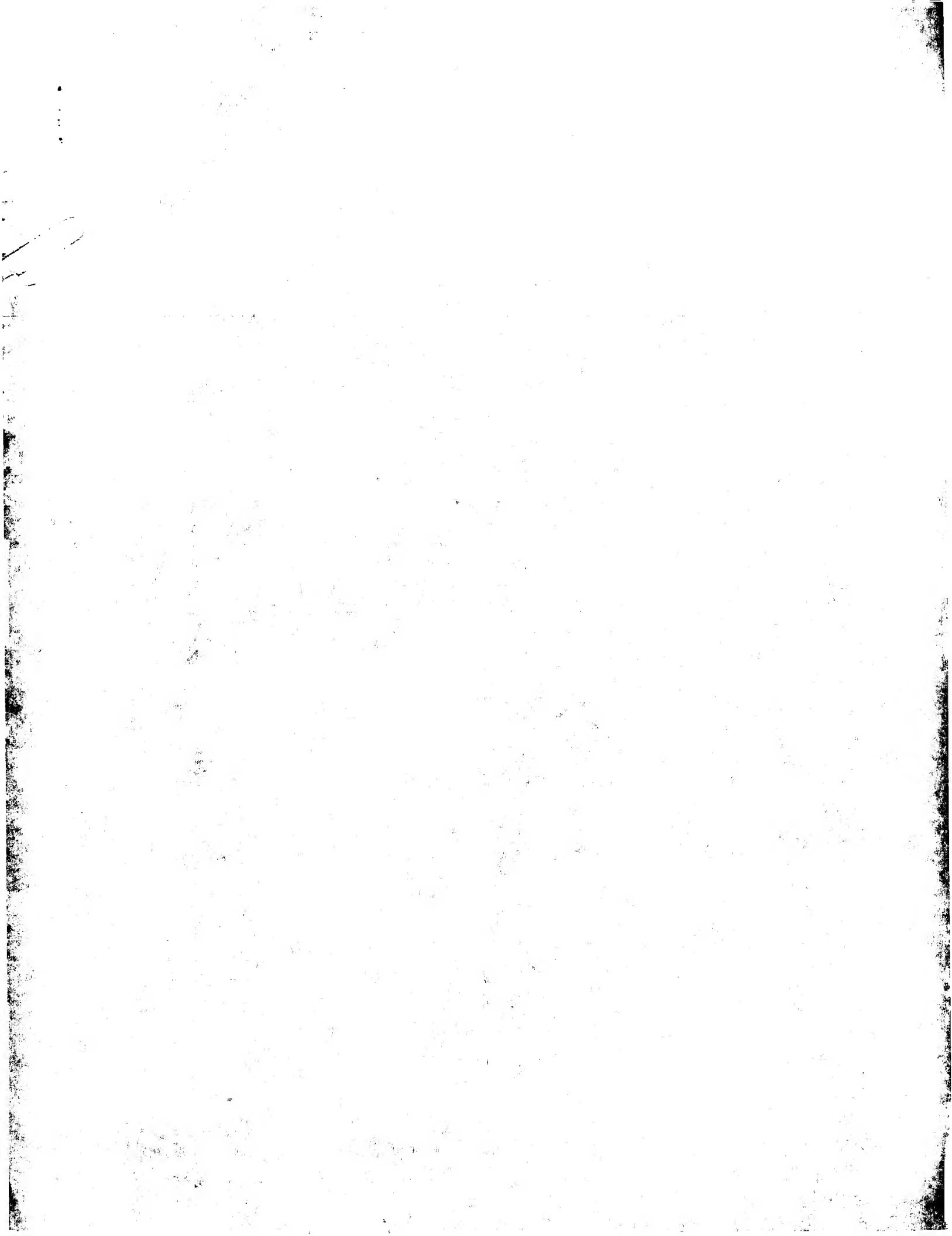
## RESULT 15

09XVZ8 PRELIMINARY; PRT; 84 AA.  
 AC 09XVZ8;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DE Y54G11A.11 PROTEIN.  
 GN Y54G11A.11.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wallis J.M.;  
 RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 investigating Biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL; AL034488; CAA22454.1; -  
 SO SEQUENCE 84 AA; 9667 MW; E996ECFC218D227C CRC64;

Query Match 100.0%; Score 20; DB 5; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VKMD 4  
 ||||  
 Db 37 VKMD 40

Search completed: October 30, 2002, 12:30:15  
 Job time : 3.24816 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:20:50 ; Search time 5.60688 Seconds  
(without alignments)  
277.344 Million cell updates/sec

Title: US-09-724-571-72

Perfect score: 63  
Sequence: 1 KTEEISEVNVXAEF 14

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	96.8	14	21	AA07888
2	60	95.2	19	22	AAE0612
3	54	85.7	18	22	AAE0609
4	54	85.7	19	22	AAE0611
5	54	85.7	20	21	AAV69714
6	54	85.7	20	21	AAV69716
7	54	85.7	30	18	AAW08360
8	54	85.7	30	20	AAI33754
9	54	85.7	30	21	AA07895
10	54	85.7	30	22	AA047264
11	54	85.7	32	17	AAW04402

12	54	85.7	32	17	AAW04403	Mouse amyloid prec
13	54	85.7	32	17	AAW04401	Mouse amyloid prec
14	54	85.7	33	18	AAW08359	Beta-secretase sub
15	54	85.7	33	20	AAI33753	Synthetic oligopep
16	54	85.7	33	21	AAW0892	Substrate for beta
17	54	85.7	39	21	AAV69718	Peptide 17-16' SW,
18	54	85.7	42	18	AAW08350	Beta-APP alpha-sec
19	54	85.7	42	20	AAI33752	Wild type APP beta
20	54	85.7	42	22	AAW07262	Synthetic oligopep
21	54	85.7	58	20	AAW98001	Peptide 26-16' SW,
22	54	85.7	115	20	AAW97997	Swedish-FAD Ape714
23	54	85.7	115	20	AAW98000	Swedish-FAD APP po
24	54	85.7	261	22	AAW0610	Chimeric cassette
25	54	85.7	506	19	AAW61152	Maltose binding pr
26	54	85.7	506	20	AAI33742	MBP-APP (SW192) fu
27	54	85.7	506	22	AAW47258	MBP-APP C-125 (Swe
28	54	85.7	695	18	AAW19504	APP695 mutant A-be
29	54	85.7	695	21	AAW8435	APP695 mutant A-be
30	54	85.7	695	22	AAW19504	Human APP695-sw va
31	54	85.7	695	22	AAW10633	Human amyloid prot
32	54	85.7	695	22	AAW06863	Human amyloid prec
33	54	85.7	695	22	AAW06607	Human amyloid prec
34	54	85.7	695	22	AAW07209	Human beta-amyloid
35	54	85.7	695	22	AAE02585	Human amyloid prec
36	54	85.7	697	21	AAW84429	Human APPSW-KK aml
37	54	85.7	697	22	AAW0636	Human amyloid prot
38	54	85.7	697	22	AAW06866	Human amyloid prec
39	54	85.7	697	22	AAW06610	Human amyloid prec
40	54	85.7	697	22	AAW07209	Human beta-amyloid
41	54	85.7	697	22	AAE02588	Human amyloid prec
42	54	85.7	751	18	AAW19505	APP751 mutant A-be
43	54	85.7	751	18	AAW19492	APP751 mutant A-be
44	54	85.7	770	18	AAW19506	APP770 mutant A-be
45	54	85.7				

#### ALIGNMENTS

RESULT 1  
ID AAB07888 standard; peptide; 14 AA.  
XX AAB07888;  
AC  
XX  
XX 14-NOV-2000 (first entry)  
DT  
XX  
DE A peptide fragment derived from beta-amyloid precursor protein.  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
PN WO200047618-A2.  
XX  
XX 17-AUG-2000.  
PD  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
XX 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX WPI; 2000-533011/48.  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.

```

PT Alzheimer's disease -
XX
XX Disclosure; Page 12; 121pp; English.
XX
XX The specification describes a beta-secretase enzyme. The enzyme cleaves
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This
CC enzyme is therefore implicated in the production of amyloid plaque
CC components which accumulate in the brains of individuals afflicted with
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to
CC a mammalian subject e.g. with Alzheimer's disease or improve cognitive
CC disease-like pathology to test if they maintain or improve cognitive
CC ability or reduce the plaque burden. The compounds are used for the
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC precursor sequence represents a peptide derived from beta-amyloid
CC precursor protein
XX
XX Sequence 14 AA:
SQ
Query Match 96.8%; Score 61; DB 21; Length 14;
Best Local Similarity 92.9%; Pred. No. 6.9e-05;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KTEETSEVNVXVAEF 14
Db 1 KTEETSEVNVXVAEF 14
|||||
|

RESULT 2
AAE00612
ID AAE00612 standard; peptide; 19 AA.
XX
XX AAE00612;
XX
XX 02-JUL-2001 (first entry)
XX
XX pCBCL APP beta-secretase cleavage site (D/I) mutant.
XX
XX Amyloid precursor protein; APP; caspase; beta-secretase; cleavage site;
XX cysteine protease; apoptosis; caspase expression cassette; metastasis;
XX tumour; cathepsin B; urokinase; proliferation; gene therapy;
XX interdomain linker; Alzheimer's disease; mutant; mutain.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX FH
XX FT Miss-difference 12 /note= "Wild-type Asp substituted by Ile"
XX
XX WO200129232-A2.
XX
XX 26-APR-2001.
XX
XX 19-OCT-2000; 2000WO-US28941.
XX
XX 20-OCT-1999; 99US-0160559.
XX PR 14-AUG-2000; 2000US-0225564.
XX
XX (SCIO-) SCIOS INC.
XX
XX Cordell B, Li Y;
XX
XX WPI; 2001-290920/30.
XX
XX Novel fusion polypeptide comprising first and second caspase subunit
XX separated by cleavage site not associated in nature with caspase
XX subunit, useful for cloning gene encoding enzymes involved in
XX proteolytic cleavage
XX
XX Disclosure; Fig 29A; 116pp; English.
XX
XX The present amino acid sequence is a pCBCL plasmid amyloid precursor
XX protein (APP) beta-secretase cleavage site (AAE00611) (D/I) mutant. APP
XX sequence containing this mutation is less susceptible to beta-secretase

```

```

CC cleavage. APP beta-secretase cleavage site is used to construct an
CC artificially engineered chimeric cassette comprising human caspase-3 with
CC interdomain linker replaced by Swedish mutant beta-secretase cleavage
CC site. This modified caspase-3 plays a pivotal role in Alzheimer's
CC disease. Caspases are a family of cysteine proteases, that participate
CC in the initiation and execution of apoptosis.
CC The present invention relates to a method for functional cloning of genes
CC encoding proteins or enzymes involved in proteolytic cleavage. The
CC invention is based on the use of caspase expression cassettes comprising
CC the coding sequence of a proteolytic cleavage site flanked by sequences
CC encoding two caspase subunits. A fusion polypeptide comprising a first
CC and a second caspase subunit, separated by a cleavage site not associated
CC in nature, is useful for cloning gene encoding enzymes involved in
CC proteolytic cleavage. An expression cassette containing fusion
CC polypeptide is used to identify a mutant cell line deficient in an
CC enzyme of interest and is also useful for diagnosis and suppression of
CC proliferation or metastases of a tumour cell characterised by
CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,
CC selectively expressed in the tumour cells). DNA encoding fusion
CC polypeptide is used in gene therapy.
XX
XX Sequence 19 AA:
SQ
Query Match 95.2%; Score 60; DB 22; Length 19;
Best Local Similarity 85.7%; Pred. No. 0.00015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KTEETSEVNVXVAEF 14
Db 2 KTEETSEVNVXVAEF 15
|||||
|

RESULT 3
AAE00609
ID AAE00609 standard; peptide; 18 AA.
XX
XX AAE00609;
XX
XX 02-JUL-2001 (first entry)
XX
XX Beta-amyloid precursor protein beta-secretase cleavage site mutant.
XX
XX Beta-amyloid precursor protein; beta-APP; caspase; beta-secretase;
XX cysteine protease; apoptosis; caspase expression cassette; metastasis;
XX tumour; cathepsin B; urokinase; proliferation; gene therapy;
XX interdomain linker; cleavage site; Alzheimer's disease; mutant; mutain.
XX
XX Unidentified.
XX
XX WO200129232-A2.
XX
XX 26-APR-2001.
XX
XX 19-OCT-2000; 2000WO-US28941.
XX
XX 20-OCT-1999; 99US-0160559.
XX PR 14-AUG-2000; 2000US-0225564.
XX
XX (SCIO-) SCIOS INC.
XX
XX Cordell B, Li Y;
XX
XX WPI; 2001-290920/30.
XX
XX Novel fusion polypeptide comprising first and second caspase subunit
XX separated by cleavage site not associated in nature with caspase
XX subunit, useful for cloning gene encoding enzymes involved in
XX proteolytic cleavage
XX
XX Example 2; Page 26; 116pp; English.
XX
XX The present sequence is a beta-secretase cleavage site of beta-amyloid
XX precursor protein (beta-APP) mutant found in certain families of

```

CC autosomal dominant form of Alzheimer's disease. This sequence is used to  
 CC construct a chimeric cassette comprising human caspase-3 with interdomain  
 CC linker replaced by this sequence. This modified caspase-3 plays a pivotal  
 CC role in Alzheimer's disease. Caspases are a family of cysteine proteases,  
 CC that participate in the initiation and execution of apoptosis.  
 CC The present invention relates to a method for functional cloning of genes  
 CC encoding proteins or enzymes involved in proteolytic cleavage. The  
 CC invention is based on the use of caspase expression cassettes comprising  
 CC the coding sequence of a proteolytic cleavage site flanked by sequences  
 CC encoding two caspase subunits. A fusion polypeptide comprising a first  
 CC and a second caspase subunit, separated by a cleavage site not associated  
 CC in nature, is useful for cloning gene encoding enzymes involved in  
 CC proteolytic cleavage. An expression cassette containing fusion  
 CC polypeptide is used to identify a mutant cell line deficient in an  
 CC enzyme of interest and is also useful for diagnosis and suppression of  
 CC proliferation or metastases of a tumour cell characterised by  
 CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.

SO Sequence 18 AA;

Query Match 85.7%; Score 54; DB 22; Length 18;  
 Best Local Similarity 85.7%; Pred. No. 0.0018;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 Db 1 KTEISEVNLDAEF 14

RESULT 4  
 AAE00611

ID AAE00611 standard; peptide; 19 AA.

AC AAE00611;

DT 02-JUL-2001 (first entry)

DE Amyloid precursor protein (APP) beta-secretase cleavage site.

KW Amyloid precursor protein; APP; caspase; beta-secretase; cleavage site;

KM cysteine protease; apoptosis; caspase expression cassette; metastasis;

KW tumour; cathepsin B; urokinase; proliferation; gene therapy;

interdomain linker; Alzheimer's disease.

OS Unidentified.

OS WO200129232-A2.

PN 26-APR-2001.

PD 19-OCT-2000; 2000MO-US28941.

PF 20-OCT-1999; 98US-0160559.

PR 14-AUG-2000; 2000US-0225564.

XX (SCIO-) SCIOS INC.

XX Cordell B, Li Y;

XX WPI: 2001-290920/30.

XX Novel fusion polypeptide comprising first and second caspase subunit

PT separated by cleavage site not associated in nature with caspase

PT subunit, useful for cloning gene encoding enzymes involved in

PT proteolytic cleavage -

XX Disclosure; Fig 28A; 116pp; English.

XX The present amino acid sequence is a beta-secretase cleavage site of  
 CC an amyloid precursor protein (APP). This sequence is used to construct  
 CC an artificially engineered chimeric cassette comprising human caspase-3

CC with interdomain linker replaced by swedish mutant beta-secretase  
 CC cleavage site. This modified caspase-3 plays a pivotal role in  
 CC Alzheimer's disease. Caspases are a family of cysteine proteases, that  
 CC participate in the initiation and execution of apoptosis.  
 CC The present invention relates to a method for functional cloning of genes  
 CC encoding proteins or enzymes involved in proteolytic cleavage. The  
 CC invention is based on the use of caspase expression cassettes comprising  
 CC the coding sequence of a proteolytic cleavage site flanked by sequences  
 CC encoding two caspase subunits. A fusion polypeptide comprising a first  
 CC and a second caspase subunit, separated by a cleavage site not associated  
 CC in nature, is useful for cloning gene encoding enzymes involved in  
 CC proteolytic cleavage. An expression cassette containing fusion  
 CC polypeptide is used to identify a mutant cell line deficient in an  
 CC enzyme of interest and is also useful for diagnosis and suppression of  
 CC proliferation or metastases of a tumour cell characterised by  
 CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.

SO Sequence 19 AA;

Query Match 85.7%; Score 54; DB 22; Length 19;  
 Best Local Similarity 85.7%; Pred. No. 0.0019;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 Db 2 KTEISEVNLDAEF 15

RESULT 5  
 AAY69714

ID AAY69714 standard; peptide; 20 AA.

AC AAY69714;

DT 11-APR-2000 (first entry)

DE Beta-APP alpha-secretase substrate [NL]-APP(-10,+10).

KW Neotropic; neuroprotective; beta-amyloid precursor protein; metabolism;

KM cleavage site; beta-secretase; neurodegenerative disease;

KW Alzheimer's disease.

OS Homo sapiens.

OS Synthetic.

PN WO9964587-A1.

PD 16-DEC-1999.

PF 04-JUN-1999; 99MO-FR01326.

PR 05-JUN-1998; 98FR-0007068.

PR 31-MAR-1999; 99US-0122599.

XX (RHON) RHONE-POULENC RORER SA.

XX (UYPA-) UNIV CURIE PARIS VI P & M.

XX Rhoiam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;

XX WPI: 2000-097537/08.

XX Polypeptide with beta-secretase activity, specific for wild-type

PT amyloid precursor protein, useful in treating Alzheimer's disease -

PS Example 3; Page 24; 44pp; French.

XX Peptides AAY69702-Y69718 represent synthetic peptide substrates for a  
 CC novel polypeptide with beta-secretase activity that can cleave  
 CC specifically the natural beta-amyloid precursor protein (bAPP). Normal  
 CC cleavage of the protein occurs between amino acids Met596-Asp597 and  
 CC Val636-Ile637 (positions 4-5 and 44-45 of AAY69701). The novel

CC polypeptide is used to identify agents that interact specifically with  
 CC it. These agents regulate metabolism of APP, particularly they slow down  
 CC or reduce production of beta-amyloid, so can be used to treat  
 CC neurodegenerative diseases, particularly Alzheimer's disease.

XX Sequence 20 AA;

Query Match 85.7%; Score 54; DB 21; Length 20;

Best Local Similarity 85.7%; Pred. No. 0.002; Mismatches 2; Indels 0; Gaps 0;

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 1 KTEISEVNDIAEF 14

RESULT 6

AAV69716  
 ID AAV69716 standard; peptide: 20 AA.

XX AAV69716;

XX 11-APR-2000 (first entry)

XX Beta-APP alpha-secretase substrate [NM]-APP(-10,+10).

XX Neurotrophic; neuroprotective; beta-amyloid precursor protein; metabolism;

KW cleavage site; beta-secretase; neurodegenerative disease;

KW Alzheimer's disease.

XX Homo sapiens.

OS Synthetic.

PN MO9964587-A1.

PD 16-DEC-1999.

PF 04-JUN-1999; 99WO-FR01326.

XX 05-JUN-1998; 98FR-0007068.

PR 31-MAR-1999; 99US-0122599.

XX (RHON ) RHONE-POULENC ROKER SA.

PA (UYPA-) UNIV CURIE PARIS VI P & M.

PI Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;

XX WPI; 2000-097537/08.

DR Polypeptide with beta-secretase activity, specific for wild-type

PT amyloid precursor protein, useful in treating Alzheimer's disease

XX Example 3; Page 24; 44pp; French.

XX Peptides AAV69702-669718 represent synthetic peptide substrates for a

CC novel polypeptide with beta-secretase activity that can cleave

CC specifically the natural beta-amyloid precursor protein (bAPP). Normal

CC cleavage of the protein occurs between amino acids Met96-Asp97 and

CC Val636-Ile637 (positions 4-5 and 44-45 of AAV69701). The novel

CC polypeptide is used to identify agents that interact specifically with

CC it. These agents regulate metabolism of APP, particularly they slow down

CC or reduce production of beta-amyloid, so can be used to treat

CC neurodegenerative diseases, particularly Alzheimer's disease.

RESULT 7  
 ID AAV08360 standard; peptide: 30 AA.

XX AAV08360;

XX 05-SEP-1997 (first entry)

XX Beta-secretase substrate #2.

XX Beta-cleavage site; beta-amyloid precursor protein; APP; beta-secretase;

KW alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.

XX Synthetic.

PN WO9640885-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US09985.

XX 07-JUN-1995; 95US-0485152.

PR 07-JUN-1995; 95US-0480498.

XX (ATHE-) ATHENA NEUROSCIENCES INC.

PA Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;

PI Mcconlogue LC, Sima S, Tan H;

XX WPI; 1997-052304/05.

DR Beta-secretase which specifically cleaves beta-amyloid precursor

XX protein; useful to screen for inhibitors useful in treatment of

PT Alzheimer's disease

XX Disclosure; Page 45; 92pp; English.

XX AAV08359-W08362 represent substrates for the enzyme of the invention.

CC The enzyme of the invention is beta-secretase, and specifically cleaves

CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP

CC is thought to occur via cleavage between residues 16 and 17 of the

CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing

CC is thought to occur by beta-secretase cleavage of beta-APP. A method of

CC beta-secretase activity can be detected and measured using a method of

CC the invention, which detects at least one of the beta-secretase cleavage

CC products formed on cleavage. The method can be used to determine whether

CC a test substance inhibits proteolytic cleavage, by beta-secretase, of

CC beta-APP. Compounds effective to at least partially inhibit

CC beta-secretase activity can be used to inhibit cleavage of beta-APP in

CC cells or mammalian hosts. Isolation and purification of beta-secretase

CC will permit chemical modelling of a critical event in the pathology of

CC Alzheimer's disease.

XX Sequence 30 AA;

Query Match 85.7%; Score 54; DB 18; Length 30;

Best Local Similarity 85.7%; Pred. No. 0.0032; Mismatches 2; Indels 0; Gaps 0;

Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 17 KTEISEVNDIAEF 30

RESULT 8

AAV33754  
 ID AAV33754 standard; protein: 30 AA.

XX AAV33754;

XX 09-NOV-1999 (first entry)

XX	Synthetic oligopeptide 26-4'SN.
DE	
XX	Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;
KM	Alzheimer's disease; measure activity; cleavage site.
XX	
OS	Synthetic.
XX	
FN	US5942400-A.
PD	
24	AUG-1999.
XX	
PE	07-JUN-1996; 96US-065984.
XX	
PR	07-JUN-1996; 96US-065984.
PR	07-JUN-1995; 95US-0480498.
XX	07-JUN-1995; 95US-0485152.
PA	(ELAN-) ELAN PHARM INC.
P1	
Anderson JP, Jacobson-Croak KL, Sinha S;	
DR	WPI: 1999-517417/43.
XX	
PT	A method for detecting human beta-secretase cleavage of polypeptides
XX	useful for identifying beta-secretase inhibitors
XX	
PS	Examples; Column 30; 43pp; English.
XX	
CC	Sequences AMY33752-Y33756 are synthetic oligopeptides used for measuring
CC	the activity of beta-secretase (AMY33741). Beta-secretase is capable of
CC	cleaving beta-amyloid protein precursor (APP) (AMY33742). These
CC	synthetic peptides contain the cleavage site of APP. Beta-secretase and
CC	APP are used in a method for detecting human beta-secretase cleavage of
CC	polypeptides and for identifying beta-secretase inhibitors. Inhibition
CC	of beta-secretase activity would be useful for chemical modelling of a
CC	critical event in the pathology of Alzheimer's disease. Inhibitors of
CC	beta-secretase would be useful for the prevention and treatment of
CC	Alzheimer's disease and Down's Syndrome.
XX	
Sequence 30 AA;	
30	

Query Match	85.7%	Score 54;	DB 20;	Length 30;
Best Local Similarity	85.7%	Pred. No. 0.0032;		
Matches 12; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	KTEETSEVNVAAEF	14	
db	17	KTEETSEVNLDAEF	30	

XX	RESULT 9
XX	AAB07895
ID	AAB07895 standard; Peptide; 30 AA.
XX	
AC	AAB07895;
XX	
DT	14-NOV-2000 (first entry)
DE	
XX	Substrate for beta-secretase enzyme.
XX	
KM	Beta-secretase: beta-amyloid precursor protein; beta-amyloid peptide
KM	amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
KM	inhibitor.
XX	
OS	Synthetic.
XX	
XX	
FH	Key
FT	Cleavage-site
XX	26..27
PN	
XX	WO200047618-A2.
XX	
XX	
DD	17-AUG-2000.

xx	10-FEB-2000; 2000MO-US03819.
pf	
xx	
xx	10-FEB-1999; 99US-0119571.
pr	15-JUN-1999; 99US-0139172.
xx	
xx	(ELAN-) ELAN PHARM INC.
pa	
pi	Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;
pi	Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
xx	
dr	WPI: 2000-533011/48.
xx	
xx	Purified beta-secretase protein used in assays to discover inhibitors
pt	which can be used for the treatment of amyloidogenic diseases e.g.
pr	Alzheimer's disease -
xx	
ps	Example 4; Page 71; 121pp; English.
xx	
cc	The specification describes a beta-secretase enzyme. The enzyme cleaves
cc	beta-amyloid precursor protein to produce beta-amyloid peptide. This
cc	enzyme is therefore implicated in the production of amyloid plaque
cc	components which accumulate in the brains of individuals afflicted with
cc	Alzheimer's disease. Inhibitors of beta-secretase are administered to
cc	a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
cc	disease-like pathology to test if they maintain or improve cognitive
cc	ability or reduce the plaque burden. The compounds are used for the
cc	treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
cc	present sequence represents a peptide substrate used to test the
cc	activity of beta-secretase enzyme.
xx	
xx	
sq	Sequence 30 AA:
Query Match	85.7%; Score 54; DB 21; Length 30;
Best Local Similarity	85.7%; Pred. No. 0.0032;
Matches 12; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	1 KTEETSEVNXVAEF 14
DB	17 KTEETSEVNDLAEF 30

RESULT 10	
AA047264	
ID	AA047264 standard; Peptide: 30 AA.
XX	
AC	AA047264;
XX	
DT	18-JUL-2001 (first entry)

DE	Peptide 26-4'SW, for used in beta-secretase assay.
XX	
KW	Beta-secretase; isotype; beta-amyloid precursor protein; APP;
KW	Beta-amyloid peptide; beta-AP; Alzheimer's disease; Downs syndrome;
KW	HCHWA-D; Swedish mutation; maltose binding protein; MBP.
XX	
OS	Synthetic.
XX	
XX	US6221645-B1.
PN	
XX	24-APR-2001.
PD	
XX	07-JUN-1996; 96US-0660531.
PF	
XX	07-JUN-1995; 95US-0480498.
PR	
XX	(ELAN-) ELAN PHARM INC.
XX	
PA	
XX	
PI	Chrysler SMS, Sinha S, Keim PS, Anderson JP, Tan H, McConlogue LC,
XX	WPI; 2001-315578/33.
DR	
XX	Novel antibody that specifically binds native beta-secretase protein.

PT	useful for raising anti-idiotypicantibodies and for detecting or
PT	diagnosing pathological conditions related to presence of respective
PT	antigens
XX	
PS	Example; Column 29; 42pp; English.
XX	
CC	The sequences given in AAB47262-67 represent synthetic peptides
CC	containing the cleavage sites derived from wild-type beta-amyloid
CC	precursor protein (APP). These peptides were used in assays utilising
CC	partially purified beta-secretase to identify beta-secretase inhibitors.
CC	Beta-secretase is thought to be responsible for the pathogenic
CC	processing of APP to form beta amyloid peptide (beta-AP) in beta-AP
CC	related conditions, e.g. Alzheimer's disease, Downs syndrome, HCHWA-D
CC	etc. Beta-secretase has a molecular weight of 260-300 kD and will bind
CC	to wheat germ agglutinin but not to concanavalin A. Beta-secretase
CC	will cleave both the wild type and the Swedish mutation of APP.
XX	
SO	Sequence 30 AA;
QY	Query Match 85.7%; Score 54; DB 22; Length 30;
Db	Best Local Similarity 85.7%; Pred. No. 0.0032;
	Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
	1 KTEETSEVNVAAEF 14
	17 KTEETSEVNVDAEF 30
RESULT 1	
AAW04402	
ID	AAW04402 standard; Protein; 32 AA.
XX	
AC	AAW04402;
XX	
DT	21-JUL-1997 (first entry)
XX	
DE	Mouse amyloid precursor protein exon 16 mutant ST59.
XX	
KW	Exon 16; murine; mouse; amyloid; precursor; protein; APP;
KW	humanisation; homozygous; heterozygous; human; Abeta; Swedish;
KW	familial; Alzheimer's; disease; PAD; mutation; tool; model;
KW	elucidation; pathology; symptomatology; screen; inhibition;
KW	transgenic; mutant; ST59.
XX	
OS	Mus spp.
XX	
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 13
FT	/note= "corresponding codon GAA"
FT	Misc-difference 15
FT	/note= "wild type Lys substituted with Asn"
FT	Misc-difference 16
FT	/note= "wild type Met substituted with Leu"
FT	Misc-difference 21
FT	/note= "wild type Gly substituted with Arg"
XX	
PN	W09634097-A1.
XX	
PD	31-OCT-1996.
XX	
PF	26-APR-1996; 96MO-US05824.
XX	
PR	23-APR-1996; 96US-0636876.
PR	26-APR-1995; 95US-0429207.
XX	
PA	(CEPH-) CEPHALON INC.
PI	Hoffman EK, Reaume AG, Scott RW, Siman R, Trusko SP;
XX	
DR	WPI, 1996-497629/49.
DR	N-PSDB; AAT38667.
XX	
PT	transgenic mice with humanised amyloid precursor protein gene -

PT	having at least 1 Swedish FAD mutation, useful as tools or models to
PT	elucidate role of human A-beta in Alzheimer's disease
XX	
XX	
PS	Example 2; Fig 12; 123pp; English.
XX	
CC	The present sequence is the protein encoded by a partial sequence
CC	from exon 16 of the mouse amyloid precursor protein (APP) gene,
CC	into which 4 humanising base pair changes have been introduced. The
CC	exon was then used in the preparation of mice homozygous or
CC	heterozygous for a targeted APP encoding gene, comprising a human
CC	A-beta peptide encoding sequence in place of the endogenous murine
CC	sequence, and at least 1 Swedish Familial Alzheimer's disease (FAD)
CC	mutation. The mice can be used as tools, or models to elucidate the
CC	role of human A-beta in AD pathology and symptomatology. They can
CC	also be used to screen chemical compounds for the ability to
CC	inhibit <i>in vivo</i> processing of APP, to yield the human A-beta peptide
CC	by administering the chemical compounds to a mouse and measuring
CC	the relative amounts of amyloidogenic and nonamyloidogenic
CC	processing of APP in a sample from the mouse at an appropriate
CC	interval after administration of the chemical compounds.
XX	
SO	Sequence 32 AA:
XX	
Query Match	85.7%; Score 54; DB 17; Length 32;
Best Local Similarity	85.7%; Fred. No. 0.0035;
Matches 12; Conservative	0; Mismatches 2; Indels 0; Gaps 0.
OY	1 KTEISEVNXVAEF 14
Db	7 KTEISEVNLDAEF 20
XX	
RESULT 12	
AAW04403	
ID	AAW04403 standard; Protein; 32 AA.
XX	
NC	AAW04403;
XX	
DT	21-JUL-1997 (first entry)
XX	
DE	Mouse amyloid precursor protein exon 16 mutant ST61.
XX	
KW	Human 16; murine; mouse; amyloid; precursor; protein; APP;
KW	humanisation; homozygous; heterozygous; human; A-beta; Swedish;
KW	familial; Alzheimer's; disease; FAD; mutation; tool; model;
KW	elucidation; pathology; symptomatology; screen; inhibition;
KW	transgenic; mutant; ST61.
XX	
OS	Mus spp.
XX	
FH	
FT	Key Location/Qualifiers
FT	Misc-difference 13
FT	/note= "corresponding codon GAA"
FT	Misc-difference 15
FT	/note= "wild type Lys substituted with Asn"
FT	Misc-difference 16
FT	/note= "wild type Met substituted with Leu"
FT	Misc-difference 26
FT	/note= "wild type Phe substituted with Tyr"
FT	Misc-difference 29
FT	/note= "wild type Arg substituted with His"
XX	
PN	W09634097-A1.
XX	
PD	31-OCT-1996.
XX	
PF	26-APR-1996; 96WO-US05824.
XX	
XX	23-APR-1996; 96GS-0636876.
PR	26-APR-1995; 95US-0429207.
XX	
PA	(CEPH-) CEPHALON INC.
XX	



PA	(CEPH-) CEPHALON INC.
PI	Hoffman EK, Reaume AG, Scott RW, Siman R, Trusko SP;
XX	
DR	WPI; 1996-497629/49.
DR	N-PSDB; AAT38666.
XX	
PT	Transgenic mice with humanised amyloid precursor protein gene -
PT	having at least 1 Swedish FAD mutation, useful as tools or models to
PT	elucidate role of human A-beta in Alzheimer's disease
XX	
PS	Example 2; Fig 12; 123pp; English.
XX	
CC	The present sequence is the protein encoded by a partial sequence
CC	from exon 16 of the mouse amyloid precursor protein (APP) gene,
CC	into which 3 humanising base pair changes have been introduced. The
CC	exon was then used in the preparation of mice homozygous or
CC	heterozygous for a targeted APP encoding gene, comprising a human
CC	Abeta peptide encoding sequence in place of the endogenous murine
CC	sequence, and at least 1 Swedish Familial Alzheimer's Disease (FAD)
CC	mutation. The mice can be used as tools, or models to elucidate the
CC	role of human Abeta in AD pathology and symptomatology. They can
CC	also be used to screen chemical compounds for the ability to
CC	inhibit in vivo processing of APP, to yield the human Abeta peptide
CC	by administering the chemical compounds to a mouse and measuring
CC	the relative amounts of amyloidogenic and nonamyloidogenic
CC	processing of APP in a sample from the mouse at an appropriate
CC	interval after administration of the chemical compounds.
XX	
SQ	Sequence 32 AA;
XX	
Query Match	85.7%; Score 54; DB 17; Length 32;
Best Local Similarity	85.7%; Pred. No. 0.0035;
Matches 12; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	1 KTEETSEVNVAAEF 14
DB	7 KTEETSEVNLDAEF 20
XX	
RESULT 14	
AAW08359	
ID	AAW08359 standard; peptide: 33 AA.
AC	AAW08359;
XX	
DT	05-SEP-1997 (first entry)
XX	
DE	Beta-secretase substrate #1.
XX	
KW	Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;
KW	alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.
XX	
OS	Synthetic.
XX	
XX	MO9640885-A2.
XX	
PD	19-DEC-1996.
XX	
FE	07-JUN-1996; 96WO-US09985.
XX	
PR	07-JUN-1995; 95US-0485152.
PR	07-JUN-1995; 95US-0480498.
XX	
FA	(ATHE-) ATHENA NEUROSCIENCES INC.
XX	
PI	Anderson JP, Chrysler SMS, Jacobson-croak KL, Kelm PS;
PI	Mcconlogue LC, Sinha S, Tan H;
XX	
DR	WPI; 1997-052304/05.
XX	
PT	Beta-secretase which specifically cleaves beta-amyloid precursor
PT	protein - useful to screen for inhibitors useful in treatment of

PT Alzheimer's disease  
 XX  
 PS Disclosure; Page 44; 92pp; English.  
 XX  
 CC AAM08359-W08362 represent substrates for the enzyme of the invention.  
 CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP  
 CC is thought to occur via cleavage between residues 16 and 17 of the  
 CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing  
 CC is thought to occur by beta-secretase cleavage of beta-APP.  
 CC Beta-secretase activity can be detected and measured using a method of  
 CC the invention, which detects at least one of the beta-secretase cleavage  
 CC products formed on cleavage. The method can be used to determine whether  
 CC a test substance inhibits proteolytic cleavage, by beta-secretase, of  
 CC beta-APP. Compounds effective to at least partially inhibit  
 CC beta-secretase activity can be used to inhibit cleavage of beta-APP in  
 CC cells or mammalian hosts. Isolation and purification of beta-secretase  
 CC will permit chemical modelling of a critical event in the pathology of  
 CC Alzheimer's disease.  
 XX  
 SQ Sequence 33 AA;  
 Query Match 85.7%; Score 54; DB 18; Length 33;  
 Best Local Similarity 85.7%; Pred. No. 0.0036;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 8 KTEEISEVNLDAEF 21  
 RESULT 15  
 AAY33753  
 ID AAY33753 standard; Protein: 33 AA.  
 XX  
 AC AAY33753;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Synthetic oligopeptide 17-16'SW.  
 XX  
 KW Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;  
 KW Alzheimer's disease; measure activity; cleavage site.  
 XX  
 OS Synthetic.  
 XX  
 PN US5942400-A.  
 XX  
 PD 24-AUG-1999.  
 XX  
 PF 07-JUN-1996; 96US-0659984.  
 XX  
 PR 07-JUN-1996; 96US-0659984.  
 PR 07-JUN-1995; 95US-0480498.  
 PR 07-JUN-1995; 95US-0485152.  
 XX  
 PA (ELAN-) ELAN PHARM INC.  
 XX  
 PI Anderson JP, Jacobson-Croak KL, Sinha S;  
 DR WPI; 1999-517417/43.  
 XX  
 PT A method for detecting human beta-secretase cleavage of polypeptides  
 PT useful for identifying beta-secretase inhibitors  
 XX  
 PS Examples; Column 30; 43pp; English.  
 XX  
 CC Sequences AAY33752-Y33756 are synthetic oligopeptides used for measuring  
 CC the activity of beta-secretase (AAY33741). Beta-secretase is capable of  
 CC cleaving beta-amyloid protein precursor (APP) (AAY33742). These  
 CC synthetic peptides contain the cleavage site of APP. Beta-secretase and  
 CC APP are used in a method for detecting human beta-secretase cleavage of  
 CC polypeptides and for identifying beta-secretase inhibitors. Inhibition

CC of beta-secretase activity would be useful for chemical modelling of a  
 CC critical event in the pathology of Alzheimer's disease. Inhibitors of  
 CC beta-secretase would be useful for the prevention and treatment of  
 CC Alzheimer's disease and Down's Syndrome.  
 XX  
 SQ Sequence 33 AA;  
 Query Match 85.7%; Score 54; DB 20; Length 33;  
 Best Local Similarity 85.7%; Pred. No. 0.0036;  
 Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 8 KTEEISEVNLDAEF 21  
 Search completed: October 30, 2002, 12:27:04  
 Job time : 7.60688 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using SW model

Run on: October 30, 2002, 12:24:20 ; Search time 1.99509 Seconds  
(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-72

Perfect score: 63  
Sequence: 1 KTEETSEVNXVAEF 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	85.7	30	2	US-08-659-984A-17
2	54	85.7	30	4	US-08-660-531-17
3	54	85.7	33	2	US-08-659-984A-16
4	54	85.7	33	4	US-08-660-531-16
5	54	85.7	42	2	US-08-659-984A-15
6	54	85.7	42	4	US-08-660-531-15
7	54	85.7	506	2	US-08-659-984A-21
8	54	85.7	506	4	US-08-660-531-21
9	54	85.7	506	4	US-09-054-334-4
10	48	76.2	58	1	US-08-371-930-25
11	48	76.2	58	1	PCT-US94-01712-25
12	48	76.2	132	6	5187153-4
13	48	76.2	162	6	5220013-4
14	48	76.2	162	6	5223482-4
15	48	76.2	264	1	US-07-990-893-5
16	48	76.2	487	1	US-08-462-859A-9
17	48	76.2	487	1	US-08-123-659A-9
18	48	76.2	487	1	US-08-464-247A-9
19	48	76.2	487	1	US-08-464-248A-9
20	48	76.2	492	1	US-08-462-859A-7
21	48	76.2	492	1	US-08-123-659A-7
22	48	76.2	492	1	US-08-464-247A-7
23	48	76.2	492	1	US-08-464-248A-7
24	48	76.2	656	1	US-08-371-930-23
25	48	76.2	656	1	PCT-US94-01712-23
26	48	76.2	676	1	US-08-371-930-24
27	48	76.2	676	5	PCT-US94-01712-24

28	48	76.2	694	1	US-08-339-152A-18	Sequence 18, Appl
29	48	76.2	694	2	US-08-007-999B-5	Sequence 3, Appl
30	48	76.2	694	2	US-08-689-276A-5	Sequence 5, Appl
31	48	76.2	695	1	US-08-371-930-25	Sequence 2, Appl
32	48	76.2	695	1	US-08-123-702-2	Sequence 30, Appl
33	48	76.2	695	1	US-08-339-152A-90	Sequence 1, Appl
34	48	76.2	695	2	US-08-104-165-1	Sequence 1, Appl
35	48	76.2	695	3	US-08-464-250-1	Sequence 1, Appl
36	48	76.2	695	4	US-08-464-250-1	Sequence 1, Appl
37	48	76.2	695	4	US-09-458-481B-4	Sequence 4, Appl
38	48	76.2	695	4	US-09-458-481B-5	Sequence 5, Appl
39	48	76.2	695	4	US-09-458-481B-6	Sequence 6, Appl
40	48	76.2	695	4	US-09-458-481B-7	Sequence 7, Appl
41	48	76.2	695	4	US-09-458-481B-8	Sequence 8, Appl
42	48	76.2	695	5	PCT-US94-01712-27	Sequence 27, Appl
43	48	76.2	695	6	5218100-2	Patent No. 5218100
44	48	76.2	751	1	US-08-123-702-4	Sequence 4, Appl
45	48	76.2	751	2	US-08-104-165-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-659-984A-17  
; Sequence 17, Application US/08659984A  
; Patent No. 5942400  
GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Sinha, Sukanto  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
; TITLE OF INVENTION: Inhibition  
; NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
; ADDRESS: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,984A  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,152  
; FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002810US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-659-984A-17  
Query Match 85.7%; Score 54; DB 2; Length 30;  
Best Local Similarity 85.7%; Pred. No. 0.00038;  
Matches 12; Conservative 0; Mismatches 2; Gaps 0;

OY 1 KTEISEVNVXAEF 14  
Db 17 KTEISEVNLDAEF 30

## RESULT 2

US-08-660-531-17  
; Sequence 17, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002210US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-660-531-17

Query Match 85.7%; Score 54; DB 4; Length 30;  
Best Local Similarity 85.7%; Pred. No. 0.00038; Indels 0; Gaps 0;  
Matches 12; Conservative 0; Mismatches 2;

OY 1 KTEISEVNVXAEF 14  
Db 17 KTEISEVNLDAEF 30

## RESULT 3

US-08-659-984A-16  
; Sequence 16, Application US/08659984A  
; Patent No. 5942400  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Jacobson-Croak, Kirsten L.  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,984A  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,152  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002810US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-659-984A-16

Query Match 85.7%; Score 54; DB 2; Length 33;  
Best Local Similarity 85.7%; Pred. No. 0.00042;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KTEISEVNVXAEF 14  
Db 8 KTEISEVNLDAEF 21

## RESULT 4

US-08-660-531-16  
; Sequence 16, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match  
Best Local Similarity 85.7%; Score 54; DB 4; Length 33;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
DB 8 KTEISEVNLDAEF 21

RESULT 5  
US-08-659-984A-15  
Sequence 15, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-15

Query Match  
Best Local Similarity 85.7%; Score 54; DB 2; Length 42;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
DB 17 KTEISEVNLDAEF 30

RESULT 6  
US-08-660-531-15  
Sequence 15, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match  
Best Local Similarity 85.7%; Score 54; DB 4; Length 42;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
DB 17 KTEISEVNLDAEF 30

RESULT 7  
US-08-659-984A-21  
Sequence 21, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2422  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-659-984A-21

Query Match 85.7%; Score 54; DB 2; Length 506;  
Best Local Similarity 85.7%; Pred. No. 0.0096;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEIIEVNVXVAEF 14  
DB 398 KTEIIEVNVLDAAEF 411

RESULT 8  
US-08-660-531-21  
Sequence 21, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-660-531-21

Query Match 85.7%; Score 54; DB 4; Length 506;  
Best Local Similarity 85.7%; Pred. No. 0.0096;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEIIEVNVXVAEF 14  
DB 398 KTEIIEVNVLDAAEF 411

RESULT 9  
US-09-054-334-4  
Sequence 4, Application US/09054334  
Patent No. 6329163  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
APPLICANT: Sinha, Sukanto  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,334  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 015270-002820US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-054-334-4

Query Match 85.7%; Score 54; DB 4; Length 506;  
Best Local Similarity 85.7%; Pred. No. 0.0096;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KTEIIEVNVXVAEF 14

Db 398 KTEISEVNMDAEF 411

## RESULT 10

US-08-371-930-25  
; Sequence 25, Application US/08371930

; Patent No. 5578451

; GENERAL INFORMATION:

; APPLICANT: Nishimoto, Ikuro

; TITLE OF INVENTION: ALZHEIMER'S DISEASE THERAPEUTICS

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 502 or 55SX

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/371,930

; FILING DATE:

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/019,208

; FILING DATE: February 18, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 00786/154001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 58

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-08-371-930-25

Query Match 76.2%; Score 48; DB 1; Length 58;  
Best Local Similarity 78.6%; Pred. No. 0.012;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNMVAAEF 14

Db 37 KTEISEVKMDAEF 50

## RESULT 11

PCT-US94-01712-25

; Sequence 25, Application PC/US9401712

; GENERAL INFORMATION:

; APPLICANT: Nishimoto, Ikuro

; TITLE OF INVENTION: ALZHEIMER'S DISEASE THERAPEUTICS

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 502 or 55SX

; OPERATING SYSTEM: MS-DOS (Version 5.0)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/01712

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/019,208

; FILING DATE: February 18, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 00786/154001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 58

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; PCT-US94-01712-25

Query Match 76.2%; Score 48; DB 5; Length 58;  
Best Local Similarity 78.6%; Pred. No. 0.012;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNMVAAEF 14

Db 37 KTEISEVKMDAEF 50

## RESULT 12

5187153-4

; Patent No. 5187153

; APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO

; TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S

; AMYLOID POLYPEPTIDE DERIVATIVES

; NUMBER OF SEQUENCES: 33

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/502,273

; FILING DATE: 29-MAR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 361,912

; FILING DATE: 06-JUN-1989

; APPLICATION NUMBER: 359,911

; FILING DATE: 12-MAY-1989

; APPLICATION NUMBER: 87,002

; FILING DATE: 18-AUG-1987

; APPLICATION NUMBER: 8,810

; FILING DATE: 30-JAN-1987

; APPLICATION NUMBER: 948,376

; FILING DATE: 31-DEC-1986

; APPLICATION NUMBER: 932,193

; FILING DATE: 17-NOV-1986

; SEQ ID NO: 4

; LENGTH: 152

; 5187153-4

Query Match 76.2%; Score 48; DB 6; Length 152;  
Best Local Similarity 78.6%; Pred. No. 0.035;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNMVAAEF 14

Db 64 KTEISEVKMDAEF 77

## RESULT 13

5220013-4  
; Patent No. 5220013

APPLICANT: PONTE, PHYLLIS A.;CORDELL, BARBARA  
TITLE OF INVENTION: DNA SEQUENCE USEFUL FOR THE DETECTION  
OF ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/444,118  
FILING DATE: 30-NOV-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 87,002  
FILING DATE: 18-AUG-1987  
APPLICATION NUMBER: 8,810  
FILING DATE: 30-JAN-1987  
APPLICATION NUMBER: 948,376  
FILING DATE: 31-DEC-1986  
APPLICATION NUMBER: 932,193  
FILING DATE: 17-NOV-1986  
SEQ ID NO:4:  
LENGTH: 162  
5220013-4

Query Match 76.2%; Score 48; DB 6; Length 162;  
Best Local Similarity 78.6%; Pred. No. 0.038;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
||||||| |||  
DB 71 KTEISEVKMDAEF 84

RESULT 14  
5223482-4  
PATENT NO. 5223482  
APPLICANT: SCHILLING, JAMES W.;PONTE, PHYLLIS A.;CORDELL,  
BARBARA  
TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE  
INHIBITORY AMYLOID PROTEIN AND METHOD OF USE  
NUMBER OF SEQUENCES: 34  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/361,912  
FILING DATE: 06-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 359,911  
FILING DATE: 12-MAY-1989  
APPLICATION NUMBER: 87,002  
FILING DATE: 18-AUG-1987  
APPLICATION NUMBER: 8,810  
FILING DATE: 30-JAN-1987  
APPLICATION NUMBER: 948,376  
FILING DATE: 31-DEC-1986  
APPLICATION NUMBER: 932,193  
FILING DATE: 17-NOV-1986  
SEQ ID NO:4:  
LENGTH: 162  
5223482-4

Query Match 76.2%; Score 48; DB 6; Length 162;  
Best Local Similarity 78.6%; Pred. No. 0.038;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
||||||| |||  
DB 71 KTEISEVKMDAEF 84

RESULT 15  
US-07-990-893-5  
Sequence 5, Application US/07990893  
Patent No. 5547841  
GENERAL INFORMATION:  
APPLICANT: Marotta, Charles A.  
APPLICANT: Zain, Sayeeda  
TITLE OF INVENTION: Genetic Sequences Coding For Alzheimer  
TITLE OF INVENTION: Amyloid From Brain

NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/990,893  
FILING DATE: 19921215  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0932.0250003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 264 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-990-893-5

Query Match 76.2%; Score 48; DB 1; Length 264;  
Best Local Similarity 78.6%; Pred. No. 0.066;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
||||||| |||  
DB 156 KTEISEVKMDAEF 169

Search completed: October 30, 2002, 12:32:29  
Job time : 2.99509 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 2.68305 Seconds  
(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-72

Perfect score: 63  
Sequence: 1 KTEISEVNVXVAEF 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	48	76.2	82	P00438	Alzheimer's disease
2	48	76.2	695	A49795	Alzheimer's disease
3	48	76.2	695	A27485	Alzheimer's disease
4	48	76.2	695	S00550	Alzheimer's disease
5	48	76.2	770	Q8H044	Alzheimer's disease
6	42	66.7	747	JH0773	Alzheimer's disease
7	39	61.9	455	S50725	hypothetical prote
8	38	60.3	316	T23930	hypothetical prote
9	38	60.3	335	A24785	hypothetical prote
10	38	60.3	1354	T13930	tripeptidyl-peptid
11	37	58.7	1185	A42404	collagen adhesin -
12	37	57.9	587	T48382	auxin-regulated pr
13	36	57.1	512	G96555	55 kDa B regulator
14	36	57.1	513	S55889	protein phosphatas
15	36	57.1	650	T22002	hypothetical prote
16	36	57.1	837	A56681	hypothetical prote
17	36	57.1	837	JM0282	penicillin amidase
18	36	57.1	867	T05389	antigen 332 - mala
19	35	55.6	184	D72745	hypothetical prote
20	35	55.6	351	JU0056	hypothetical prote
21	35	55.6	433	C94619	flagellin - Serrat
22	35	55.6	433	C94619	probable serine ca
23	35	55.6	456	S19212	adenylosuccinate 1
24	35	55.6	456	C90829	adenylosuccinate 1
25	35	55.6	456	A85687	adenylosuccinate 1
26	35	55.6	504	AD1615	glycine betaine/ca
27	35	55.6	505	E95246	4-alpha-glucanotra
28	35	55.6	505	B98111	4-alpha-glucanotra
29	35	55.6	572	S72249	trithorax protein
30	35	55.6	731	T19721	hypothetical prote

30	35	55.6	802	2	S49252	penicillin amidase
31	35	55.6	802	2	I39665	penicillin amidase
32	34	54.0	136	2	C81436	biopolymer transpo
33	34	54.0	207	2	A47615	colanic acid capsu
34	34	54.0	225	2	E97244	colanic acid capsu
35	34	54.0	274	2	B96991	colanic acid capsu
36	34	54.0	294	2	D97693	citrate lyase, bet
37	34	54.0	294	2	AC2919	citrate lyase, bet
38	34	54.0	307	2	S38152	citrate lyase, bet
39	34	54.0	319	2	C97720	hypothetical prote
40	34	54.0	341	2	F71042	mip protein (impor
41	34	54.0	342	2	C75172	hypothetical prote
42	34	54.0	357	2	E71123	enolase related (e
43	34	54.0	366	2	B87446	probable sugar tra
44	34	54.0	412	2	C38351	hypothetical prote
45	34	54.0	415	2	T12974	phosphoprotein pho

## ALIGNMENTS

## RESULT 1

P00438

Alzheimer's disease amyloid A4 protein precursor - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 30-Sep-1993 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995

C:Accession: P00438; C60045

R:Davidson, J.S.; West, R.L.; Kotikalapudi, P.; Maroun, L.E.

Biochem. Biophys. Res. Commun. 188, 905-911, 1992

A>Title: Sequence and methylation in the beta/A4 region of the rabbit amyloid precurs

A:Reference number: P00438; MUID:93075180

A:Accession: P00438

A:Molecule type: DNA

A:Residues: 1-82 <DAV>

A:Cross-references: GB:M83558; GB:M83657

R:Johnstone, E.M.; Chaney, M.O.; Norris, F.H.; Pascual, R.; Little, S.P.

Brain Res. Mol. Brain Res. 10, 299-305, 1991

A>Title: Conservation of the sequence of the Alzheimer's disease amyloid peptide in d

A:Reference number: A60045; MUID:92017079

A:Accession: C60045

A:Molecule type: mRNA

A:Residues: 12-68 <JOH>

A:Cross-references: EMBL:X56129

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas

C:Keywords: alternative splicing; Alzheimer's disease; amyloid; Down's syndrome

Query Match 76.2%, Score 48; DB 2; Length 82;

Best Local Similarity 76.6%; Pred. No. 0.05;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KTEISEVNVXVAEF 14  
Db 7 KTEISEVNVXVAEF 20

## RESULT 2

A49795

Alzheimer's disease amyloid beta protein precursor - crab-eating macaque

C:Species: Macaca fascicularis (crab-eating macaque)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: A49795

R:Podlasky, M.B.; Tolan, D.R.; Selkoe, D.J.

Am. J. Pathol. 138, 1423-1435, 1991

A>Title: Homology of the amyloid beta protein precursor in monkey and human supports

A:Reference number: A49795; MUID:91273117

A:Accession: A49795

A:Molecule type: preliminary

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-695 <POD>

A:Cross-references: GB:M58727; NID:g342062; PIDN:AAA36829.1; PID:g342063

C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas

C:Keywords: alternative splicing

Query Match 76.2%; Score 48; DB 1; Length 695;  
 Best Local Similarity 78.6%; Pred. No. 0.55;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 587 KTEISEVKMDAEF 600

RESULT 3  
 A27485  
 Alzheimer's disease amyloid beta/A4 protein homolog precursor - mouse  
 N:Alternate names: proteinase nexin II  
 C:Species: Mus musculus (house mouse)  
 C>Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 13-Aug-1999  
 C:Accession: A27485; S19727; I49485  
 R:Yamada, T.; Sasaki, H.; Furuya, H.; Miyata, T.; Goto, I.; Sakaki, Y.  
 Biochem. Biophys. Res. Commun. 149, 665-671, 1987  
 A:Title: Complementary DNA for the mouse homolog of the human amyloid beta protein precursor  
 A:Reference number: A27485; MUID:88106489  
 A:Accession: A27485  
 A:Molecule type: mRNA  
 A:Residues: 1-695 <YAM>  
 A:Cross-references: GB:M18373; NID:g191568; PIDN:AAA37139.1; PID:g309085  
 A:Experimental source: brain  
 R:de Strooper, B.; van Leuven, F.; van den Berghe, H.  
 Biochim. Biophys. Acta 1129, 141-143, 1991  
 A:Title: The amyloid beta protein precursor or proteinase nexin II from mouse is closer  
 A:Reference number: S19727; MUID:92096458  
 A:Accession: S19727  
 A:Molecule type: mRNA  
 A:Residues: 1-210, 'G', 212-220, 'S', 222-396, 'A', 398-402, 'T', 404-448, 'A', 450-695 <STR>  
 A:Cross-references: EMBL:X59379  
 R:Iizumi, R.; Yamada, T.; Yoshikai, S.; Sasaki, H.; Hattori, M.; Sakaki, Y.  
 Gene 112, 189-195, 1992  
 A:Title: Positive and negative regulatory elements for the expression of the Alzheimer's  
 A:Reference number: I49485; MUID:92209998  
 A:Accession: I49485  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-19 <RES>  
 A:Cross-references: GB:D10603; NID:g220328; PIDN:BA01456.1; PID:g220329  
 C:Genetics:  
 A:Map position: 16C3  
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase 1  
 C:Keywords: alternative splicing; amyloid; transmembrane protein

Query Match 76.2%; Score 48; DB 2; Length 695;  
 Best Local Similarity 78.6%; Pred. No. 0.55;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 587 KTEISEVKMDAEF 600

RESULT 4  
 S00550  
 Alzheimer's disease amyloid beta protein precursor - rat  
 N:Alternate names: beta-A4 amyloid protein  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 13-Aug-1999  
 C:Accession: S00550; A11245; A39820; S46251  
 R:Shivers, B.D.; Hilbich, C.; Multhaup, G.; Salbaum, M.; Beyreuther, K.; Seeburg, P.H.  
 EMBO J. 7, 1365-1370, 1988  
 A:Title: Alzheimer's disease amyloidogenic glycoprotein: expression pattern in rat brain  
 A:Reference number: S00550; MUID:88312583  
 A:Accession: S00550  
 A:Molecule type: mRNA  
 A:Residues: 1-695 <SHI>  
 A:Cross-references: EMBL:X07648; NID:g55616; PIDN:CAA30488.1; PID:g55617  
 R:Schubert, D.; Schroeder, R.; LaCordiere, M.; Salton, T.; Cole, G.  
 Science 241, 223-226, 1988

A:Title: Amyloid beta protein precursor is possibly a heparan sulfate proteoglycan co  
 A:Reference number: A41245; MUID:88264430  
 A:Accession: A41245  
 A:Molecule type: protein  
 A:Residues: 18-37, 'X', 39-40, 'X', 42-44 <SCH>  
 A:Note: evidence for heparan sulfate attachment  
 R:Hesse, L.; Beher, D.; Masters, C.L.; Multhaup, G.  
 FEBS Lett. 349, 109-116, 1994  
 A:Title: The beta-A4 amyloid precursor protein binding to copper.  
 A:Reference number: S46251; MUID:94320627  
 A:Contents: annotation; copper binding sites  
 A:Note: rat peptides were isolated but not sequenced  
 R:Potempska, A.; Styles, J.; Mehta, P.; Kim, K.S.; Miller, D.L.  
 J. Biol. Chem. 266, 8464-8469, 1991  
 A:Title: Purification and tissue level of the beta-amyloid peptide precursor of rat b  
 A:Reference number: A39820; MUID:9121087  
 A:Accession: A39820  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 18-32 <POT>  
 A:Experimental source: brain  
 C:Comment: Deposition of amyloid protein as neurofibrillary tangles and/or plaques is  
 C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinas  
 C:Keywords: alternative splicing; amyloid; glycoprotein; transmembrane protein  
 F:625-648/Domain: transmembrane #status predicted <YAM>

Query Match 76.2%; Score 48; DB 2; Length 695;  
 Best Local Similarity 78.6%; Pred. No. 0.55;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 587 KTEISEVKMDAEF 600

RESULT 5  
 OHRD4  
 Alzheimer's disease amyloid beta protein precursor [validated] - human  
 N:Alternate names: Alzheimer's disease amyloid A4 protein; coagulation factor Xla inh  
 N:Contains: amyloid beta protein long, plaque form; amyloid beta protein short, vascu  
 protein precursor splice form APP(770)  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Jun-1987 #sequence\_revision 28-Jul-1995 #text\_change 15-Sep-2000  
 C:Accession: S02260; S05194; A32277; A33260; A33486; I39452; I39451; I39453; I59562;  
 4688; A29583; A29302; A60805; J10038; S06121; A38384; S29076; S38252;  
 R:Lemaire, H.G.; Salbaum, J.M.; Multhaup, G.; Kang, J.; Bayney, R.M.; Unterbeck, A.;  
 Nucleic Acids Res. 17, 517-522, 1989  
 A:Title: The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid is encode  
 A:Reference number: S02260; MUID:89128427  
 A:Accession: S02260  
 A:Molecule type: DNA  
 A:Residues: 1-288, 'V', 365-770 <LEM1>  
 A:Cross-references: EMBL:X13466  
 A:Note: alternative splice form APP(695)  
 R:Lemaire, H.G.  
 submitted to the EMBL Data Library, November 1988  
 A:Reference number: S05194  
 A:Accession: S05194  
 A:Molecule type: DNA  
 A:Residues: 1-14, 'VW', 17-288, 'V', 365-770 <LEM2>  
 A:Cross-references: EMBL:X13466; NID:g35598; PIDN:CAA31830.1; PID:g871360  
 A:Note: alternative splice form APP(695)  
 R:La Faut, G.; Lahiri, D.K.; Salton, S.R.J.; Robakis, N.K.  
 Biochem. Biophys. Res. Commun. 159, 297-304, 1989  
 A:Title: Characterization of the 5'-end region and the first two exons of the beta-pr  
 A:Reference number: A32277; MUID:89165870  
 A:Accession: A32277  
 A:Molecule type: DNA  
 A:Residues: 1-75 <LAF>  
 A:Cross-references: GB:M24546; GB:M24547; NID:g341202; PIDN:AC13654.1; PID:g516074  
 R:Johnstone, E.M.; Chaney, M.O.; Moore, R.E.; Ward, K.E.; Norris, F.H.; Little, S.P.  
 Biochem. Biophys. Res. Commun. 163, 1248-1255, 1989  
 A:Title: Alzheimer's disease amyloid peptide is encoded by two exons and shows simila

A:Reference number: A33260; MUID:89392030  
A:Accession: A33260  
A:Molecule type: DNA  
A:Residues: 656-737 <JOH>  
A:Cross-references: GB:M29270; NID:g178663; PIDN:AA51768.1; PID:g178665  
R:Prelli, F.; Levy, E.; van Duijn, S.G.; Bots, G.T.A.M.; Luyendijk, W.; Frangione, B.  
Biochem. Biophys. Res. Commun. 170, 301-307, 1990  
A:Title: Expression of a normal and variant Alzheimer's beta-protein gene in amyloid of  
A:Reference number: A35486; MUID:90321244  
A:Accession: A35486  
A:Molecule type: DNA  
A:Residues: 672-710 <PRE1>  
A:Note: 693-gln was found in DNA isolated from HCMA-D patients  
R:Yoshikuni, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sasaki, Y.  
Gene 87, 257-263, 1990  
A:Title: Genomic organization of the human amyloid beta-protein precursor gene.  
A:Reference number: 139451; MUID:90236318  
A:Accession: 139451  
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME  
A:Molecule type: DNA  
A:Residues: 1-770 <YOS1>  
A:Cross-references: GB:M3112; NID:g178613; PIDN:AA59502.1; PID:g178616  
A:Accession: 139451  
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME  
A:Molecule type: DNA  
A:Residues: 1-530, QWLMPTVPAFWAKVGR' <YOS2>  
A:Cross-references: GB:M34875; NID:g178608; PIDN:AA59501.1; PID:g178615  
R:Yoshikuni, S.I.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sasaki, Y.  
Gene 102, 291-292, 1991  
A:Reference number: A59020; MUID:91340168  
A:Accession: A59020  
A:Contents: annotation: erratum  
A:Note: revised physical map for reference 139451  
R:Levy, E.; Carman, M.D.; Fernandez-Madrid, I.J.; Power, M.D.; Lieberburg, I.; van Duine  
Science 248, 1124-1126, 1990  
A:Title: Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrh  
A:Reference number: 139453; MUID:90260663  
A:Accession: 139453  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 656-737 <LEV>  
A:Cross-references: GB:M37896; NID:g178618; PIDN:AA5127.1; PID:g178620  
A:Note: a mutation with 693-gln is presented  
R:Murrell, J.; Farlow, M.; Ghetti, B.; Benson, M.D.  
Science 254, 97-99, 1991  
A:Title: A mutation in the amyloid precursor protein associated with hereditary Alzheimer  
A:Reference number: 139562; MUID:9202553  
A:Accession: 139562  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 689-716, 'F', 718-737 <MGR>  
A:Cross-references: GB:S57665; NID:g236720; PIDN:AA19991.1; PID:g236721  
R:Kamino, K.; Orr, H.T.; Payami, H.; Wajsbom, E.M.; Alonso, M.E.; Palst, S.M.; Anderson,  
ataki, S.E.; Korenberg, J.R.; Sharma, V.; Kukull, W.; Larson, E.; Heston, L.L.; Martin,  
Am. J. Hum. Genet. 51, 998-1014, 1992  
A:Title: Linkage and mutational analysis of familial Alzheimer disease kindreds for the  
A:Reference number: AA4017; MUID:93035397  
A:Accession: AA4017  
A:Molecule type: DNA  
A:Residues: 687-692, G', 694-718 <KAM1>  
A:Cross-references: GB:S45136; NID:g257379; PIDN:AA23646.1; PID:g257380  
A:Experimental source: familial Alzheimer disease family 1IT  
A:Note: sequence extracted from NCBI backbone (NCBI:115376)  
R:Kang, J.; Lemaire, H.G.; Unterbeck, A.; Salbaum, J.M.; Masters, C.L.; Grzeschik, K.H.;  
Nature 325, 733-736, 1987  
A:Title: The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface  
A:Reference number: A03134; MUID:87144572

A:Accession: A03134  
A:Molecule type: mRNA  
A:Residues: 1-288, 'V', 365-770 <KAN>  
A:Cross-references: GB:Y00264; NID:g28525; PIDN:CA68374.1; PID:g28526  
A:Note: alternative splice form APP(695)  
R:Robakis, N.K.; Ramakrishna, N.; Wolfe, G.; Wisniewski, H.M.  
Proc. Natl. Acad. Sci. U.S.A. 84, 4190-4194, 1987  
A:Title: Molecular cloning and characterization of a cDNA encoding the cerebrovascula  
A:Reference number: A29030; MUID:87231971  
A:Accession: A29030  
A:Molecule type: mRNA  
A:Residues: 284-288, 'V', 365-646, 'E', 648-770 <ROB>  
A:Cross-references: GB:M16765; NID:g178539; PIDN:AA51722.1; PID:g178540  
R:Goldgaber, D.; Herman, M.I.; McBride, O.W.; Saffioti, U.; Gajdusek, D.C.  
Science 235, 877-880, 1987  
A:Title: Characterization and chromosomal localization of a cDNA encoding brain amylo  
A:Reference number: AA7584; MUID:87120328  
A:Accession: AA7584  
A:Molecule type: mRNA  
A:Residues: 674-756, 'S', 758-770 <GOL>  
A:Cross-references: GB:M1533; NID:g178706; PIDN:AA35540.1; PID:g178707  
A:Experimental source: brain  
R:Tanzi, R.E.; Gusella, J.F.; Watkins, P.C.; Bruns, G.A.P.; St George-Hyslop, P.; Van  
Science 235, 880-884, 1987  
A:Title: Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near  
A:Reference number: AA7585; MUID:87120329  
A:Accession: AA7585  
A:Molecule type: mRNA  
A:Residues: 674-703 <TAN1>  
A:Cross-references: GB:M15332; NID:g177957; PIDN:AA51564.1; PID:g177958  
R:DiYcks, T.; Weidemann, A.; Multhaup, G.; Salbaum, J.M.; Lemaire, H.G.; Kang, J.; Mue  
EMBO J. 7, 949-957, 1988  
A:Title: Identification, transmembrane orientation and biogenesis of the amyloid A4 p  
A:Reference number: S02638; MUID:88296437  
A:Accession: S02638  
A:Molecule type: mRNA  
A:Residues: 672-678 <DYR>  
R:Tanzi, R.E.; McClatchey, A.I.; Lamert, E.D.; Villa-Komaroff, L.; Gusella, J.F.; N  
Nature 331, 528-530, 1988  
A:Title: Protease inhibitor domain encoded by an amyloid protein precursor mRNA assoc  
A:Reference number: S00707; MUID:88122640  
A:Accession: S00707  
A:Molecule type: mRNA  
A:Residues: 286-344, 'I', 365-366 <TAN2>  
A:Cross-references: EMBL:X06982; NID:g28817; PIDN:CA30042.1; PID:g28812  
A:Experimental source: promyelocytic leukemia cell line HL60  
R:Ponte, P.; Gonzalez-Demhilt, P.; Schilling, J.; Miller, J.; Hsu, D.; Greenberg, B.;  
Nature 331, 525-527, 1988  
A:Title: A new A4 amyloid mRNA contains a domain homologous to serine proteinase inh  
A:Reference number: S00925; MUID:88122639  
A:Accession: S00925  
A:Molecule type: mRNA  
A:Residues: 1-344, 'I', 365-770 <PO2>  
A:Cross-references: GB:X06989; EMBL:Y00297; NID:g28720; PIDN:CA30050.1; PID:g28721  
A:Note: alternative splice form APP(751)  
R:Kitaguchi, N.; Takahashi, Y.; Tokushima, Y.; Shiojiri, S.; Ito, H.  
Nature 331, 530-532, 1988  
A:Title: Novel precursor of Alzheimer's disease amyloid protein shows protease inh  
A:Reference number: A38949; MUID:88122641  
A:Accession: A38949  
A:Molecule type: mRNA  
A:Residues: 287-367 <KIT>  
A:Cross-references: GB:X06981; NID:g28816; PIDN:CA30041.1; PID:g28811  
A:Experimental source: glioblastoma cell line  
A:Note: alternative splice form APP(770)  
R:Vitek, M.P.; Rasool, C.G.; de Sauvage, F.; Vitek, S.M.; Bartus, R.T.; Beer, B.; Ash  
Brain Res. Mol. Brain Res. 4, 121-131, 1988  
A:Title: Absence of mutation in the beta-amyloid cDNAs cloned from the brains of thre  
A:Reference number: A30320  
A:Accession: A30320  
A:Status: not compared with conceptual translation

```

A:Molecule type: mRNA
A:Residues: 284-288, 'V', 365-770 <YIT1>
A:Accession: B30320
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 122-288, 'V', 365-770 <YIT2>
A:Accession: C30320
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 506-770 <YIT3>
R:Zain, S.B., Salim, M., Chou, W.G., Sajdel-Sulowska, E.M., Majocha, R.E., Marotta, C.A.
Proc. Natl. Acad. Sci. U.S.A. 85, 929-933, 1988
A:Title: Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer disease br
A:Reference number: A31087; MUID:88124954
A:Accession: A31087
A:Molecule type: mRNA
A:Residues: 507-770 <ZAI>
A:Cross-references: GB:M8734; NID:g178572; PIDN:AAA51726.1; PID:g178573
A:Note: the authors translated the codon GAA for residue 599 as Gly, ACC for residue 603
8 as Val, GTG for residue 609 as Asn, AAT for residue 610 as Gly, and GGT for residue 65
A:Note: the cited Genbank accession number, J03594, is not in release 101.0
R:Masters, C.L., Multhaup, G., Simms, G., Pottgiesser, J., Martins, R.N., Beyreuther, K.

Query Match          76.2%; Score 48; DB 1; Length 770;
Best Local Similarity 78.6%; Pred. No. 0.62;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14
|||||||
Db 662 KTEISEVKMAEF 675

RESULT 6
JH0773
Alzheimer's disease amyloid beta protein precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 13-Aug-1999
C:Accession: JH0773
R:Okado, H., Okamoto, H.,
Biochem. Biophys. Res. Commun. 189, 1561-1568, 1992
A:Title: A Xenopus homologue of the human beta-amyloid precursor protein: developmental
A:Reference number: JH0773; MUID:93129227
A:Accession: JH0773
A:Molecule type: mRNA
A:Residues: 1-747 <OKA>
A:Cross-references: GB:S52417; NID:9263150; PIDN:AAB24853.1; PID:9263151
A:Experimental source: larva
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteinase
C:Keywords: alternative splicing; amyloid
F:287-337/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Query Match          66.7%; Score 42; DB 2; Length 747;
Best Local Similarity 64.3%; Pred. No. 8.3;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14
|||||||
Db 639 KTEISEVKMDSY 652

RESULT 7
S50725
hypothetical protein YNL206c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N1346
C:Species: Saccharomyces cerevisiae
C>Date: 23-Aug-1995 #sequence_revision 19-Jan-1996 #text_change 21-Jul-2000
C:Accession: S50725; S63163; S49863
R:Joniaux, J.L., Coster, F., Purnelle, B., Goffeau, A.
Yeast 10, 1639-1645, 1994
A:Title: A 21.7 kb DNA segment on the left arm of yeast chromosome XIV carries WH13, GCH
A:Reference number: S50712; MUID:95242839
A:Accession: S50725
A:Status: nucleic acid sequence not shown

```

```

A:Molecule type: DNA
A:Residues: 1455 <JON>
A:Cross-references: EMBL:X78898; NID:g600045; PIDN:CAA55502.1; PID:g600059
R:Coster, F., Joniaux, J.L., Goffeau, A., Purnelle, B., Van Dyck, L.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63151
A:Accession: S63163
A:Molecule type: DNA
A:Residues: 1-455 <COS>
A:Cross-references: EMBL:Z71482; NID:g1302206; PID:e239612; PID:g1302207; MIPS:YNL206
A:Experimental source: strain S288C
C:Genetics:

A:Map position: 14L

Query Match          61.9%; Score 39; DB 2; Length 455;
Best Local Similarity 63.6%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KTEISEVNXV 11
|||||
Db 63 KTEISEINTI 73

RESULT 8
T23930
hypothetical protein R05D7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23930
R:Dobson, R.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19819
A:Accession: T23930
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-316 <WIL>
A:Cross-references: EMBL:Z81105; PIDN:CAR03217.1; GSPDB:GN00019; CESP:R05D7.3
A:Experimental source: clone R05D7
C:Genetics:
A:Gene: CESP:R05D7.3
A:Map position: 1
A:Introns: 63/3; 97/3; 122/3; 195/2; 230/1

Query Match          60.3%; Score 38; DB 2; Length 316;
Best Local Similarity 46.2%; Pred. No. 18;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAE 13
|||||
Db 283 ETEITEISEISE 295

RESULT 9
A24785
hypothetical protein 335 - slime mold (Dictyostelium discoideum) transposon DIRS-1
C:Species: Dictyostelium discoideum
C>Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 18-Jun-1993
C:Accession: A24785
R:Cappello, J., Handelsman, K., Lodish, H.F.
Cell 43, 105-115, 1985
A:Title: Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted
A:Reference number: A94654; MUID:86079481
A:Accession: A24785
A:Molecule type: DNA
A:Residues: 1-335 <CAP>

Query Match          60.3%; Score 38; DB 2; Length 335;
Best Local Similarity 57.1%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14
|||||
Db 115 KXDEISELNKVFNF 128

```

```
RESULT 10
113930
tripeptidyl-peptidase II (EC 3.4.14.10) - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13930
R:Renn, S.C.P.; Tomkinson, B.; Taghert, P.H.
J. Biol. Chem. 273, 19173-19182, 1998
A:Title: Characterization and cloning of tripeptidyl peptidase II from the fruit fly, D.
A:Reference number: 217814; MUID:98334655
A:Accession: T13930
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1354 <REN>
A:Cross-references: EMBL:AF035251; NID:g3387807; PID:g3387808; PIDN:AAC28563.1
C:Genetics:
A:Map position: 2
C:Keywords: dipeptidylpeptide hydrolase

Query Match
Best Local Similarity 60.3%; Score 38; DB 2; Length 1354;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 EISEVNXVAEF 14
Db 929 EKISEANLVASF 940

RESULT 11
A42404
collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C:Accession: A42404; S27665
R:Petit, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wlberg, K.; Lindberg, M.; Hook, M.
J. Biol. Chem. 267, 4766-4772, 1992
A:Title: Molecular characterization and expression of a gene encoding a staphylococcus a
A:Reference number: A42404; MUID:92165839
A:Contents: FDA 574
A:Accession: A42404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1185 <PAT>
A:Cross-references: EMBL:M81736
A:Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match
Best Local Similarity 58.7%; Score 37; DB 2; Length 1185;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14
Db 118 KVEKLSVSGFAEF 131

RESULT 12
T48582
auxin-regulated protein GH3 homolog T31B5.170 - Arabidopsis thaliana
N:Alternate names: protein T31B5.170
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 18-Jul-2001
C:Accession: T48582
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24490
A:Accession: T48582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-587 <BEV>
A:Cross-references: EMBL:AL163491
```

```
A:Experimental source: cultivar Columbia; BAC clone T31B5
C:Genetics:
A:Map position: 5
A:Introns: 97/2; 131/2; 178/2; 400/1
A:Note: T31B5.170
C:Superfamily: soybean auxin-regulated protein GH3

Query Match
Best Local Similarity 57.9%; Score 36.5; DB 2; Length 587;
Matches 10; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

OY 1 KTEE---ISEVNXVAEF 14
Db 562 KTEEAVKILEANVSEF 578

RESULT 13
G96555
55 kDa B regulatory subunit of phosphatase 2A [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96555
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G96555
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:AE005173; NID:g11094761; PIDN:AAG29694.1; GSPDB:GN00141
C:Genetics:
A:Gene: F19C24.10
A:Map position: 1

Query Match
Best Local Similarity 57.1%; Score 36; DB 2; Length 512;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14
Db 277 KMEDLSEVITSAEF 290

RESULT 14
S55889
protein phosphatase 2A B regulatory chain 55K - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S55889
R:Rundell, S.J.; Hartung, A.J.; Corum III, J.W.; O'Neill, M.
Plant Mol. Biol. 28, 257-266, 1995
A:Title: Characterization of a cDNA encoding the 55 kDa B regulatory subunit of Arabi
A:Reference number: S55889; MUID:95322588
A:Accession: S55889
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-513 <RUN>
A:Cross-references: GB:U18129; NID:g7103729; PIDN:AAA86695.1; PID:g710330

Query Match
Best Local Similarity 57.1%; Score 36; DB 2; Length 513;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14
```

Db 278 KMEIIEVITSAEF 291

## RESULT 15

T22002  
hypothetical protein F39H11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T22002

R:White, S.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19500

A:Accession: T22002

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-650 <WILL>

A:Cross-references: EMBL:Z81079; PIDN:CAH03084.1; GSPDB:GN00019; CESP:F39H11.4

A:Experimental source: clone F39H11

C:Genetics:

A:Gene: CESP:F39H11.4

A:Map position: 1

A:Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 57.1%; Score 36; DB 2; Length 650;

Best local similarity 42.9%; Pred. No. 99;

Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEIIEVNYVAEF 14

||:| ||: ||:|

Db 122 KQESSEIQRISDF 135

Search completed: October 30, 2002, 12:31:27  
Job time : 5.68305 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:10 ; Search time 1.20393 seconds

(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-72

Perfect score: 63

Sequence: 1 KTEETSEVNVXAEF 14

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## SUMMARIES

Prod. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	48	76.2	751	1 A4_SAIISC	Q95241 salmirt sci
2	48	76.2	770	1 A4_HUMAN	P05067 homo sapien
3	48	76.2	770	1 A4_MOUSE	P12023 mus musculu
4	48	76.2	770	1 A4_RAT	P08592 rattus norv
5	39	61.9	455	1 YNU6_YEAST	P40161 saccharomyc
6	37	58.7	309	1 FSTL_FLABI	P52838 flaveria bi
7	37	58.7	351	1 CNA_STAAU	O53654 staphylococ
8	35	55.6	456	1 FLIC_SERMA	P13713 escherichia
9	35	55.6	473	1 PUR8_ECOLI	P25739 escherichia
10	35	55.6	505	1 MDM2_XENLA	P38273 xenopus lae
11	35	55.6	802	1 MALO_STIRN	P29851 streptococc
12	35	55.6	802	1 PAC_ARTVI	P31956 arthrobacte
13	35	55.6	802	1 PAC_BACME	O60136 bacillus me
14	34	54.0	207	1 RCSA_KLEAE	P56155 saccharomyc
15	34	54.0	307	1 YK55_YEAST	P36155 saccharomyc
16	34	54.0	413	1 ZABB_RABIT	Q00006 o serine/th
17	34	54.0	426	1 ZABA_PIG	Q29090 s serine/th
18	34	54.0	427	1 CISTY_ECOLI	P00891 escherichia
19	34	54.0	427	1 CISTY_SALTY	O68883 salmonella
20	34	54.0	443	1 ZABB_HUMAN	O00005 homo sapien
21	34	54.0	443	1 ZABB_PIG	P54614 sus scrofa
22	34	54.0	447	1 ZABB_RAT	P36877 r serine/th
23	34	54.0	447	1 ZABA_HUMAN	Q00007 h serine/th
24	34	54.0	447	1 ZABA_RAT	P36876 r serine/th
25	34	54.0	453	1 ZABD_RAT	P56932 r serine/th
26	34	54.0	499	1 ZABA_DROME	P36872 drosophila
27	34	54.0	620	1 Y870_METUA	O58280 methanococc
28	34	54.0	759	1 TEAL_YEAST	P47988 saccharomyc
29	34	54.0	967	1 CAP2_YEAST	P51059 zea mays (m
30	34	54.0	1064	1 Y108_METUA	O60307 methanococc
31	34	54.0	1612	1 RRPO_ORSEC	P56659 odontogloss
32	34	54.0	1613	1 VIT2_CAEEL	P55155 caenorhabdi
33	34	54.0	1616	1 VIT1_CAEEL	P55155 caenorhabdi

34	33	52.4	169	1 CHE1_ANTRO	P05686 anthracia p
35	33	52.4	234	1 BI0D_BACSH	P22818 bacillus sp
36	33	52.4	401	1 OPO2_RICR	Q92094 rickettsia
37	33	52.4	423	1 CISTY_ACTAN	P20902 actinobact
38	33	52.4	452	1 YDUE_ECOLI	P38055 escherichia
39	33	52.4	901	1 PIP_LACLA	P49022 lactococcus
40	33	52.4	956	1 PODK_FLAPR	O42736 flaveria pr
41	33	52.4	1324	1 MSH6_ARATH	O04716 arabidopsis
42	32	50.8	94	1 STXE_BUTUJ	P56637 buthocus ju
43	32	50.8	97	1 STXE_METUA	Q57802 methanococc
44	32	50.8	197	1 ENGB_ARCFU	O28943 archaeoglob
45	32	50.8	216	1 RCSB_ECOLI	P14374 escherichia

## ALIGNMENTS

```

RESULT 1
A4_SAIISC STANDARD: PRT: 751 AA.
AC Q95241:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alzheimer's disease amyloid A4 protein precursor [Contains: Beta-
DE amyloid protein (Beta-Ap) (A-beta)].
GN App.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX NCBI_TaxID=9521;
RN [1]
RP TISSUE=Liver and Kidney;
RC MEDLINE=96108492; PubMed=8532114;
RX Levy E., Amorim A., Frangione B., Walker L.C.;
RT "Beta-amyloid precursor protein gene in squirrel monkeys with
RT cerebral amyloid angiopathy.";
RL Neurobiol. Aging 16:805-808(1995).
CC -!- FUNCTION: FUNCTIONAL NEURONAL RECEPTOR WHICH COUPLES TO
CC INTRACELLULAR SIGNALING PATHWAY THROUGH THE GTP-BINDING PROTEIN
CC G(O).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION
CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC
CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE
CC NXYX MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF
CC PHOSPHORYLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: S81024; AAD14347.1; -.
CC HSSP: P05067; IAP.
CC InterPro: IPR001868; A4_APP.
CC InterPro: IPR002223; Kunitz_BPTI.
CC Pfam: PF02177; A4_EXTRA; 1.
CC Pfam: PF00014; Kunitz_BPTI; 1.
CC PRINTS: PR00203; AMYLOIDA4.
CC PRINTS: PR00759; BASICPTASE.
CC SMART: SM00131; KU; 1.
CC SMART: SM00319; A4_EXTRA; 1.
CC PROSITE: PS00320; A4_INTRA; 1.
CC PROSITE: PS00280; BPTI_KUNITZ_1; 1.
CC PROSITE: PS00279; BPTI_KUNITZ_2; 1.

```

KW glycoprotein; Amyloid; Neurone; Transmembrane; Alternative splicing;  
 KW Signal; Serine protease inhibitor.  
 FT CHAIN 1 17 BY SIMILARITY.  
 FT CHAIN 18 751 A4 PROTEIN.  
 FT CHAIN 653 695 BETA-AMYLOID PROTEIN (POTENTIAL).  
 FT CHAIN 680 680 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 18 704 POTENTIAL.  
 FT TRANSMEM 681 751 CYTOSOLASMIC (POTENTIAL).  
 FT DOMAIN 705 751 BPT/KUNITZ INHIBITOR.  
 FT SITE 287 345 CLATHRIN-BINDING (BY SIMILARITY).  
 FT SITE 740 743 REACTIVE BIND.  
 FT ACT SITE 301 302 BY SIMILARITY.  
 FT DISULFID 291 341 BY SIMILARITY.  
 FT DISULFID 300 324 BY SIMILARITY.  
 FT DISULFID 316 337 BY SIMILARITY.  
 FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (PROBABLE).  
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (PROBABLE).  
 SQ SEQUENCE 751 AA; 84893 MW; 6C3E431089569049 CRC64;

Query Match 76.28; Score 48; DB 1; Length 751;  
 Best Local Similarity 78.64; Pred. No. 0.35;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 Db 643 KTEISEVKMDAEF 656

RESULT 2  
 A4\_HUMAN STANDARD; PRT; 770 AA.  
 ID A4\_HUMAN P05067; P09000; Q16011;  
 AC 13-AUG-1987 (Rel. 05, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein precursor (Protease nexin-II)  
 DE (PN-II) (APPI) [Contains: Beta-amyloid protein (Beta-ApP) (A-beta)].  
 GN APP OR A4 OR CVAP OR AD1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA MEDLINE=87144572; PubMed=2881207;  
 RA Kang J., Lemaire H.-G., Unterbeck A., Salbaum J.M., Masters C.L.,  
 RA Grzeschik K.-H., Multhaup G., Beyreuther K., Mueller-Hill B.;  
 RA "The precursor of Alzheimer's disease amyloid A4 protein resembles a  
 RA cell-surface receptor.";  
 RT Nature 325:733-736(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=88126359; PubMed=2893289;  
 RA Ponte P., Gonzalez-Dewhilt P., Schilling J., Miller J., Hsu D.,  
 RA Greenberg B., Davis K., Wallace W., Lieberburg I., Fuller F.,  
 RA Cordell B.;  
 RA "A new A4 amyloid mRNA contains a domain homologous to serine  
 RA proteinase inhibitors.";  
 RT Nature 331:525-527(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=89128427; PubMed=2783775;  
 RA Lemaire H.-G., Salbaum J.M., Multhaup G., Kang J., Bayney R.M.,  
 RA Unterbeck A., Beyreuther K., Mueller-Hill B.;  
 RA "The PreA4(695) precursor protein of Alzheimer's disease A4 amyloid  
 RA is encoded by 16 exons.";  
 RT Nucleic Acids Res. 17:517-522(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97263807; PubMed=9108164;  
 RA Hattori M., Tsukahara F., Furuhata Y., Tanahashi H., Hirose M.,  
 RA Saito M., Tsukuni S., Sakaki Y.;  
 RA "A novel method for making nested deletions and its application for  
 sequencing of a 300 kb region of human APP locus.";  
 Nucleic Acids Res. 25:1802-1808(1997).  
 RN [5]  
 RP SEQUENCE OF 286-345 AND 365-366 FROM N.A.  
 RA MEDLINE=88122640; PubMed=2893290;  
 RA Tanzi R.E., McClatchey A.I., Lampetti E.D., Villa-Komaroff L.,  
 RA Gusella J.F., Neve R.L.;  
 RA "Protease inhibitor domain encoded by an amyloid protein precursor  
 RA mRNA associated with Alzheimer's disease.";  
 RT Nature 331:528-530(1988).  
 RN [6]  
 RP SEQUENCE OF 287-367 FROM N.A.  
 RA MEDLINE=88122641; PubMed=2893291;  
 RA Kitaguchi N., Takahashi Y., Tokushima Y., Shiojiri S., Ito H.;  
 RA "Novel precursor of Alzheimer's disease amyloid protein shows  
 RA protease inhibitory activity.";  
 RT Nature 331:530-532(1988).  
 RN [7]  
 RP SEQUENCE OF 284-289 AND 365-770 FROM N.A.  
 RA MEDLINE=87231971; PubMed=3035574;  
 RA Robakis N.K., Ramakrishna N., Wolfe G., Wisniewski H.M.;  
 RA "Molecular cloning and characterization of a cDNA encoding the  
 RA cerebrovascular and the neuritic plaque amyloid peptides.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 84:4190-4194(1987).  
 RN [8]  
 RP SEQUENCE OF 507-770 FROM N.A.  
 RA MEDLINE=88124954; PubMed=2893379;  
 RA Zain S.B., Salim M., Chou W.G., Saidel-Sulkowska E.M., Majocha R.E.,  
 RA Marotta C.A.;  
 RA "Molecular cloning of amyloid cDNA derived from mRNA of the Alzheimer  
 RA disease brain: coding and noncoding regions of the fetal precursor  
 RA mRNA are expressed in the cortex.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 85:929-933(1988).  
 RN [9]  
 RP SEQUENCE OF 672-681.  
 RA MEDLINE=88035004; PubMed=3312495;  
 RA Pardridge W.M., Vinters H.V., Yang J., Eisenberg J., Choi T.B.,  
 RA Tourtelotte W.W., Huebner V., Shively J.E.;  
 RA "Amyloid angiopathy of Alzheimer's disease: amino acid composition  
 RA and partial sequence of a 4,200-dalton peptide isolated from cortical  
 RA microvessels.";  
 RT J. Neurochem. 49:1394-1401(1987).  
 RN [10]  
 RP SEQUENCE OF 739-770 FROM N.A.  
 RA MEDLINE=90236318; PubMed=2110105;  
 RA Yoshikawa S.-I., Sasaki H., Don-Ura K., Furuya H., Sakaki Y.;  
 RA "Genomic organization of the human amyloid beta-protein precursor  
 RA gene.";  
 RT Gene 87:257-263(1990).  
 RN [11]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC TISSUE=Liver;  
 RA MEDLINE=89016647; PubMed=3140222;  
 RA Schon E.A., Mita S., Sadlock J., Herbert J.;  
 RA "A cDNA specifying the human amyloid beta precursor protein (ABPP)  
 RA encodes a 95-kDa polypeptide.";  
 RT Nucleic Acids Res. 16:9351-9351(1988).  
 RN [12]  
 RP SEQUENCE OF 18-50.  
 RA MEDLINE=87250462; PubMed=3597385;  
 RA van Nostrand W.E., Cunningham D.D.;  
 RA "Purification of protease nexin II from human fibroblasts.";  
 RT J. Biol. Chem. 263:8508-8514(1987).  
 RN [13]  
 RP IDENTITY OF APP WITH NEXIN-II.  
 RA MEDLINE=89384866; PubMed=2506449;  
 RA Oltersdorf T., Fritz L.C., Schenk D.B., Lieberburg I., Dovey H.F.,  
 RA Johnson-Wood K.L., Beattie E.C., Ward P.J., Blacher R.W., Dovey H.F.,  
 RA Sinha S.;  
 RA "The secreted form of the Alzheimer's amyloid precursor protein with  
 RA the Kunitz domain is protease nexin-II.";  
 RT Nature 341:144-147(1989).  
 RN [14]



RP PROTEASE-SPECIFICITY OF INHIBITOR DOMAIN.  
 RX MEDLINE-90211252; PubMed-1969731;  
 RA Kido H., Fukutomi A., Schilling J., Wang Y., Cordell B., Katunuma N.;  
 RT "Protease-specificity of Kunitz inhibitor domain of Alzheimer's  
 RT disease amyloid protein precursor.";  
 RL Biochem. Biophys. Res. Commun. 167:716-721(1990).  
 RN [15]  
 RP COMPLEX WITH G(O).  
 RX MEDLINE-9318965; PubMed-8446172;  
 RA Nishimoto I., Okamoto T., Matsura Y., Takahashi S., Okamoto T.,  
 RA Mayama Y., Ogata E.;  
 RT "Alzheimer amyloid protein precursor complexes with brain GTP-binding  
 RT protein G(O).";  
 RL Nature 362:75-79(1993).  
 RN [16]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 28-133.  
 RX MEDLINE-99215582; PubMed-10201399;  
 RA Rosjohn J., Cappai R., Fell S.C., Henry A., McKinstry W.J.,  
 RA Galatis D., Heese L., Multhaup G., Beyreuther K., Masters C.L.,  
 RA Parker M.W.;  
 RT "Crystal structure of the N-terminal, growth factor-like domain of  
 RT Alzheimer amyloid precursor protein.";  
 RL Nat. Struct. Biol. 6:327-331(1999).  
 RN [17]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 287-344.  
 RX MEDLINE-91104913; PubMed-2125487;  
 RA Hynes T.R., Randal M., Kennedy L.A., Eigenbrodt C., Kosiakof A.A.;  
 RT "X-ray crystal structure of the protease inhibitor domain of  
 RT Alzheimer's amyloid beta-protein precursor.";  
 RL Biochemistry 29:10018-10022(1990).  
 RN [18]  
 RP STRUCTURE BY NMR OF 289-344.  
 RX MEDLINE-9203148; PubMed-1718421;  
 RA Head S.L., Tilton R.F., Jr., Hammond L.S., Lee A., Bayney R.M.,  
 RA Kamarek M.E., Ramabhadran T.V., Dreyer R.N., Davis G., Unterbeck A.,  
 RA Tamburini P.P.;  
 RT "Sequential NMR resonance assignment and structure determination of  
 RT the Kunitz-type inhibitor domain of the Alzheimer's beta-amyloid  
 RT precursor protein.";  
 RL Biochemistry 30:10467-10478(1991).  
 RN [19]  
 RP STRUCTURE BY NMR OF 672-699.  
 RX MEDLINE-94281210; PubMed-7516706;  
 RA Talafous J., Marchowski K.J., Klopman G., Zagorski M.G.;  
 RT "Solution structure of residues 1-28 of the amyloid beta-peptide.";  
 RL Biochemistry 33:7788-7796(1994).  
 RN [20]  
 RP STRUCTURE BY NMR OF 696-706.  
 RX MEDLINE-97128622; PubMed-8973180;  
 RA Kohno T., Kobayashi K., Maeda T., Sato K., Takashima A.;  
 RT "Three-dimensional structures of the amyloid beta peptide (25-35) in  
 RT membrane-mimicking environment.";  
 RL Biochemistry 35:16094-16104(1996).  
 RN [21]  
 RP STRUCTURE BY NMR OF 672-711.  
 RX MEDLINE-98359783; PubMed-9693002;  
 RA Coles M., Blacknell W., Watson A.A., Fairlie D.P., Craik D.J.;  
 RT "Solution structure of amyloid beta-peptide(1-40) in a water-micelle  
 RT environment. Is the membrane-spanning domain where we think it is?";  
 RL Biochemistry 37:11064-11077(1998).  
 RN [22]  
 RP STRUCTURE BY NMR OF 672-699.  
 RX MEDLINE-20400066; PubMed-10940222;  
 RA Poulsen S.-A., Watson A.A., Craik D.J.;  
 RT "Solution structures in aqueous SDS micelles of two amyloid beta  
 RT peptides of Abeta(1-28) mutated at the alpha-secretase cleavage  
 RT site.";  
 RL J. Struct. Biol. 130:142-152(2000).  
 RN [23]  
 RP STRUCTURE BY NMR OF 681-706.  
 RX MEDLINE-20400065; PubMed-10940221;  
 RA Zhang S., Iwata K., Lachemann M.J., Peng J.W., Li S., Stimson E.R.,  
 RA Lu Y., Felix A.M., Maggio J.E., Lee J.P.;

RT "The Alzheimer's peptide a beta adopts a collapsed coil structure in  
 RT water.";  
 RL J. Struct. Biol. 130:130-141(2000).  
 RN [24]  
 RP SIGNAL SEQUENCE CLEAVAGE SITE, AND TOPOLOGY.  
 RX MEDLINE-88296437; PubMed-2800137;  
 RA Dykes T., Weidemann A., Multhaup G., Salbaum J.M., Lemaire H.-G.,  
 RA Kang J., Mueller-Hill B., Masters C.L., Beyreuther K.;  
 RT "Identification, transmembrane orientation and biogenesis of the  
 RT amyloid A4 precursor of Alzheimer's disease.";  
 QY 1 KTEEISEVNXVAEF 14  
 DB 662 KTEEISEVXMDAEF 675  
 Query Match 76.2%; Score 48; DB 1; Length 770;  
 Best Local Similarity 78.6%; Pred. NO. 0.36;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 RESULT 3  
 A4\_MOUSE STANDARD; PRT; 770 AA.  
 ID A4\_MOUSE  
 AC P12023;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein homolog precursor  
 DE (Amyloidogenic glycoprotein) (AG).  
 GN App.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId:10090;  
 RN [1]  
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RX MEDLINE-92096458; PubMed-1756177;  
 RA de Strooper B., van Leuven F., van den Bergh H.;  
 RT "The amyloid beta protein precursor or proteinase nexin II from mouse  
 RT is closer related to its human homolog than previously reported.";  
 RL Biochim. Biophys. Acta 1129:141-143(1991).  
 RN [2]  
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.  
 RC TISSUE-Brain;  
 RX MEDLINE-88106489; PubMed-3322280;  
 RA Yamada T., Sasaki H., Furuya H., Miyata T., Goto I., Sasaki Y.;  
 RT "Complementary DNA for the mouse homolog of the human amyloid beta  
 RT protein precursor.";  
 RL Biochem. Biophys. Res. Commun. 149:665-671(1987).  
 RN [3]  
 RP REVISIONS.  
 RA Yamada T.;  
 RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 289-364 FROM N.A.  
 RC STRAIN-CD-1; TISSUE-Placenta;  
 RX MEDLINE-89345111; PubMed-2569710;  
 RA Fukuchi K., Martin G.M., Deeb S.S.;  
 RT "Sequence of the protease inhibitor domain of the A4 amyloid protein  
 RT precursor of Mus domesticus.";  
 RL Nucleic Acids Res. 17:5396-5396(1989).  
 RN [5]  
 RP SEQUENCE OF 1-19 FROM N.A.  
 RX MEDLINE-9209998; PubMed-1555768;  
 RA Irumi R., Yamada T., Yoshikai S.I., Sasaki H., Hattori M.,  
 RA Sakai Y.;  
 RT "Positive and negative regulatory elements for the expression of the  
 RT Alzheimer's disease amyloid precursor-encoding gene in mouse.";  
 RL Gene 112:189-195(1992).  
 RN [6]  
 RP SEQUENCE OF 281-380 FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE-Brain, and Kidney;

RA MEDLINE-89149813; PubMed-2493250.  
 RA Yamada T., Sasaki H., Dohura K., Goto I., Sakaki Y.,  
 RT "Structure and expression of the alternatively-spliced forms of mRNA  
 RT for the mouse homolog of Alzheimer's disease amyloid beta protein  
 RT precursor."  
 RL Biochem. Biophys. Res. Commun. 158:906-912(1989).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS; APP(395), APP(563),  
 CC APP(751) AND APP(770) (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- TISSUE SPECIFICITY: AAA(770) IS EXPRESSED IN KIDNEY, AAA(751) IS  
 CC WIDELY EXPRESSED. AAA(695) IS EXPRESSED IN BRAIN, KIDNEY AND  
 CC LIVER.  
 CC -1- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION  
 CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC  
 CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE  
 CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF  
 CC PHOSPHORYLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BPPI/KUNITZ INHIBITOR DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X59379; -; NOT ANNOTATED\_CDS.  
 CC EMBL: M18373; AAA37139.1; -;  
 CC EMBL: X15210; CAA33280.1; -;  
 CC EMBL: D10603; BAA01456.1; -;  
 CC EMBL: M24397; AAA39929.1; -;  
 CC PIR: A27485; A27485.  
 CC PIR: S04855; S04855.  
 CC PIR: S19727; S19727.  
 CC HSP: P05067; 10CM.  
 CC MGD: MGI:88059; APP.  
 CC InterPro: IPR001868; A4\_APP.  
 CC InterPro: IPR002223; Kunitz\_BPPI.  
 CC Pfam: PF02177; A4\_EXTRA; 1.  
 CC PRINTS: PR00203; AMYLOIDA4.  
 CC PRINTS: PR00759; BASICPTASE.  
 CC SMART: SM00006; A4\_EXTRA; 1.  
 CC SMART: SM00131; KU; 1.  
 CC PROSITE: PS00319; A4\_EXTRA; 1.  
 CC PROSITE: PS00320; A4\_INTRA; 1.  
 CC PROSITE: PS00280; BPPI\_KUNITZ\_2; 1.  
 CC PROSITE: PS0279; BPPI\_KUNITZ\_2; 1.  
 CC GlycoProtex: Amyloid; Neurone; Transmembrane; Signal;  
 CC KW Alternative splicing; Serine protease inhibitor.  
 CC SIGNL 1  
 CC CHAIN 18 770  
 FT DOMAIN 18 699  
 FT TRANSMEM 700 723  
 FT DOMAIN 724 770  
 FT DOMAIN 673 715  
 FT DOMAIN 287 345  
 FT SITE 759 762  
 FT DISUFLD 291 341  
 FT DISUFLD 300 324  
 FT DISUFLD 316 337  
 FT CARBOHYD 542 542  
 FT CARBOHYD 571 571  
 FT VARSPLIC 289 289  
 FT VARSPLIC 290 364  
 FT VARSPLIC 346 380  
 FT SEQUENCE 770 AA; 86752 MW; 26C5DDE890CAFA CRC64;  
 Query Match 76.2%; Score 48; DB 1; Length 770;

Best Local Similarity 78.6%; Pred. No. 0.36;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 KTEISEVNVKVAEP 14  
 Db 662 KTEISEVNVKVAEP 675  
 RESULT 4  
 ID A4\_RAT STANDARD; PRT; 770 AA.  
 AC P08592;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alzheimer's disease amyloid A4 protein homolog precursor  
 DE (Amyloidogenic glycoprotein) (A6).  
 GN APP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 1-289 AND 365-770 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE-88312583; PubMed-2900758;  
 RA Shivers B.D., Hlilich C., Multhaup G., Salbaum J.M., Beyreuther K.,  
 RA Seeburg P.H.;  
 RT "Alzheimer's disease amyloidogenic glycoprotein: expression pattern  
 RT in rat brain suggests a role in cell contact."  
 RL EMBO J. 7:1365-1370(1988).  
 RN [2]  
 RP SEQUENCE OF 289-364 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE-89183625; PubMed-2648331;  
 RA Kang J., Mueller-Hill B.;  
 RT "The sequence of the two extra exons in rat preA4."  
 RL Nucleic Acids Res. 17:2130-2130(1989).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS; APP(395), APP(563), APP(695),  
 CC APP(751) AND APP(770) (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- DOMAIN: THE CLATHRIN-BINDING SITE IS ESSENTIAL FOR ITS ASSOCIATION  
 CC WITH X11-ALPHA, -BETA, AND -GAMMA. THE SEQUENCE SPECIFIC  
 CC RECOGNITION EXTENDS TO PEPTIDE RESIDUES THAT ARE C-TERMINAL TO THE  
 CC NPXY MOTIF. THIS INTERACTION APPEARS TO BE INDEPENDENT OF  
 CC PHOSPHORYLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BPPI/KUNITZ INHIBITOR DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X07648; CAA30488.1; -;  
 CC EMBL: X14066; CAA32229.1; -;  
 CC PIR: S00550; S00550.  
 CC PIR: S03607; S03607.  
 CC HSP: P05067; 1AAP.  
 CC InterPro: IPR001868; A4\_APP.  
 CC InterPro: IPR002223; Kunitz\_BPPI.  
 CC Pfam: PF02177; A4\_EXTRA; 1.  
 CC Pfam: PF00014; Kunitz\_BPPI; 1.  
 CC PRINTS: PR00203; AMYLOIDA4.  
 CC PRINTS: PR00759; BASICPTASE.  
 CC SMART: SM00006; A4\_EXTRA; 1.  
 CC SMART: SM00131; KU; 1.  
 CC PROSITE: PS00319; A4\_EXTRA; 1.  
 CC PROSITE: PS00320; A4\_INTRA; 1.

DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Glycoprotein; Amyloid; Neurone; Transmembrane; Signal;  
 KW Alternative splicing; Serine protease inhibitor.  
 FT SIGNAL 1 17  
 FT CHAIN 18 770  
 FT DOMAIN 18 699  
 FT TRANSMEM 700 723  
 FT DOMAIN 724 770  
 FT DOMAIN 673 715  
 FT DOMAIN 287 345  
 FT SITE 759 762  
 FT DISULFID 291 341  
 FT DISULFID 300 324  
 FT DISULFID 316 337  
 FT CARBOHYD 542 542  
 FT CARBOHYD 571 571  
 FT VARSPLIC 289 289  
 FT VARSPLIC 290 364  
 SQ SEQUENCE 770 AA; 86704 MW; C26C9D6BB2D929A7 CRC64;

Query Match 76.2%; Score 48; DB 1; Length 770;  
 Best Local Similarity 78.6%; Pred. No. 0.36;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 662 KTEEISEVXMDAEF 675

RESULT 5  
 YN06\_YEAST STANDARD; PRT; 455 AA.  
 ID YN06\_YEAST  
 AC P40161;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 51.6 kDa protein in SSB2-SPX18 intergenic region.  
 GN YN1206C OR N1346.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=95242839; PubMed=7725799;  
 RA Joniaux J.-L., Coster F., Purnelle B., Goffeau A.,  
 RT "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV  
 carries WHI3, GCR2, SPX18, SPX19, an homologue to the heat shock gene  
 SSB1 and 8 new open reading frames of unknown function."  
 RL Yeast 10.1639-1645(1994).  
 CC -1- SIMILARITY: SOME, TO DROSOPHILA AND MAMMALIAN SINGLE-STRAND  
 RECOGNITION PROTEINS (SSRP).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X78898; CAAS5502.1; -;  
 DR EMBL: Z71482; CA96106.1; -;  
 DR SCD: S0005150; YNL206C.  
 KW Hypothetical protein.

KW SEQUENCE 455 AA; 51621 MW; 79CACAC659A7F4FAD CRC64;

Query Match 61.9%; Score 39; DB 1; Length 455;  
 Best Local Similarity 63.6%; Pred. No. 9.2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KTEEISEVNXV 11  
 |||||  
 DB 63 KTEEISEVNTI 73

RESULT 6  
 FSTL\_FLABI STANDARD; PRT; 309 AA.  
 ID FSTL\_FLABI  
 AC P52838;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Flavonol sulfoltransferase-like (EC 2.8.2.-).  
 OS Flaveria bidentis.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;  
 CC Heliantheae; Flaveria.  
 OX NCBI\_TaxID=4224;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=95232186; PubMed=7716232;  
 RA Ananvoranich S., Gulick P., Ibrahim R.K.;  
 RT "Flavonol sulfoltransferase-like cDNA clone from Flaveria bidentis.";  
 RL Plant Physiol. 107:1019-1020(1995).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential)  
 CC -1- SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASES FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U10277; AAA87399.1; -;  
 DR HSSP: P50224; ICM.  
 DR InterPro: IPR000863; Sulfoltransferase.  
 DR Pfam: PF00685; Sulfoltransfer; 1.  
 DR Prodom: PD001218; Sulfoltransfer; 1.  
 KW Transferase.  
 FT BINDING 138 153  
 SQ SEQUENCE 309 AA; 35917 MW; 598A9C9EFEBB75303 CRC64;

Query Match 58.7%; Score 37; DB 1; Length 309;  
 Best Local Similarity 57.1%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 KTEEISEVNXVAEF 14  
 |||||  
 DB 211 KSEPSINVKRLAEF 224

RESULT 7  
 ID CNA\_STRAU STANDARD; PRT; 1183 AA.  
 AC Q53654;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen adhesin precursor.  
 DE CNA.  
 OS Staphylococcus aureus.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=FDA 574;  
 RX MEDLINE=92165839; PubMed=1311320;  
 RA Patti J.W., Jonsson H., Guss B., Switalski L.M., Wilberg K.,

```

RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin."
RT J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RA Erratum.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RA J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN-PDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pomoeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin."
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81736; AAA20874.1; -.
CC DR PDB: 1AMX; 24-JUN-98.
CC DR InterPro: IPR001899; Gram_pos_anchor.
CC DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
CC KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
CC FT SIGNAL 1 29
CC FT CHAIN 30 1183 POTENTIAL.
CC FT DOMAIN 30 1157 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 1158 1177 MEMBRANE ANCHOR (POTENTIAL).
CC FT DOMAIN 1178 1183 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 151 318 COLLAGEN-BINDING.
CC FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.
CC FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).
CC FT DOMAIN 1151 1156 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC FT REPEAT 533 719 BL.
CC FT REPEAT 720 906 B2.
CC FT REPEAT 907 1093 B3.
CC SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match 58.7%; Score 37; DB 1; Length 1183;
Best Local Similarity 50.0%; Pred. No. 58;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Flagellin.
GN FLAG OR FLAF OR HAG.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=274;
RX MEDLINE=89378766; PubMed=2673930;
RA Harshey R.M., Estepa G., Yanagi H.;
RT "Cloning and nucleotide sequence of a flagellin-coding gene (hag)
RT from Serratia marcescens 274."
RL Gene 79:1-8(1989)
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M27219; AAA26556.1; -.
CC DR PIR: J00056; J00056.
CC DR InterPro: IPR001029; Flagellin_C.
CC DR InterPro: IPR001492; Flagellin_N.
CC DR Pfam: PF00700; Flagellin_C_1.
CC DR Pfam: PF00669; Flagellin_N_1.
CC DR PRINTS: PR00207; FLAGELLIN.
CC DR ProDom: PD000316; Flagellin_C_1.
CC KW Flagella.
CC SQ SEQUENCE 351 AA; 36865 MW; FDA4809AEB7D3CB8 CRC64;

Query Match 55.6%; Score 35; DB 1; Length 351;
Best Local Similarity 41.7%; Pred. No. 38;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

RESULT 8
ID *FLIC_SERMA STANDARD; PRT; 351 AA.
AC P13713;

```

```

RESULT 9
ID *PUB8_ECOLI STANDARD; PRT; 456 AA.
AC P25739;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Adenylosuccinate lyase (EC 4.3.2.2) (Adenylosuccinase) (ASL).
GN PUBR OR B1131.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
RC STRAIN=K12;
RX MEDLINE=92104952; PubMed=1729205;
RA He B., Smith J.M., Zalkin H.;
RT "Escherichia coli pubR gene: cloning, nucleotide sequence, and
RT regulation by purR."
RL J. Bacteriol. 174:130-136(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```

RA Green S.M., Drabble W.T.;  
RL Submitted (May-1991) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-97061202; PubMed-8905232;  
RA Oshima T., Alta H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishio H., Nishio Y., Saito N.,  
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiiuchi T.;  
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
CC -1- CATALYTIC ACTIVITY: 1-(5'-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXYAMIDE)  
-5-AMINOIMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-  
IMIDAZOLECARBOXYAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =  
FUMARATE + AMP).  
CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE LYASE I FAMILY. ADENYLOSUCINATE LYASE  
SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M74924; AAA92731.1; -;  
DR EMBL: X59307; CAA41996.1; -;  
DR EMBL: AE000213; AAC74215.1; -;  
DR EMBL: D90748; BAA35953.1; -;  
DR EMBL: D90749; BAA35962.1; -;  
DR PIR: A43307; A43307.  
DR PIR: S19212; S19212.  
DR HSSP: P02633; 1BOD.  
DR Ecocore: EGI1314; 1BOD.  
DR InterPro: IPR000362; Fumarate-lyase.  
DR Pfam: PF00206; lyase\_1; 1.  
DR PROSITE: PS00163; FUMARATE\_LYASES; 1.  
KW Purine biosynthesis; Lyase; Complete proteome.  
FT ACT\_SITE 91 91 ACID (BY SIMILARITY).  
FT ACT\_SITE 171 171 BASE (BY SIMILARITY).  
FT CONFLICT 145 145 P -> A (IN REF. 1).  
FT CONFLICT 154 154 I -> L (IN REF. 1).  
SO SEQUENCE 456 AA; 51542 MW; 8D1FA546B6795BC CRC64;  
  
Query Match 55.6%; Score 35; DB 1; Length 456;  
Best Local Similarity 50.0%; Pred. No. 50;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 KTEIISVYNXVAEF 14  
Db 103 KVAEIPELHAVSEF 116  
ID MDM2\_XENLA STANDARD; PRT; 473 AA.  
AC P56273;

DT 15-JUL-1998 (rel. 36, Created)  
DT 30-MAY-2000 (rel. 39, Last sequence update)  
DT 01-MAR-2002 (rel. 41, Last annotation update)  
DE Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (P53-binding protein  
Mdm2) (Double minute 2 protein) (Xdm2).  
GN MDM2.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-97282706; PubMed-9136986;  
RA Marchal V., Elenbaas B., Taneyhill L., Piette J., Mechali M.,  
RA Nicolas J.-C., Levine A.J., Moreau J.;  
RT "Conservation of structural domains and biochemical activities of the  
MDM2 protein from *Xenopus laevis*.";  
RL Oncogene 14:1427-1433(1997).  
[2]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 13-119 IN COMPLEX WITH P53.  
RX MEDLINE-97081050; PubMed-8675929;  
RA Kussie P.H., Gorina S., Marchal V., Elenbaas B., Moreau J.,  
RA Levine A.J., Pavletich N.P.;  
RT "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor  
transactivation domain.";  
RL Science 274:948-953(1996).  
CC -1- FUNCTION: MAY BIND P53 PROTEIN AND MAY FUNCTION AS AN UBIQUITIN  
LIGASE E3.  
CC -1- COFACTOR: ZINC IS REQUIRED FOR UBIQUITIN LIGASE E3 ACTIVITY (BY  
SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).  
CC -1- DEVELOPMENTAL STAGE: EXPRESSION INCREASES FROM OOCYTE STAGE I/II  
TO REACH ITS MAXIMUM IN OOCYTE STAGE V/VI IN UNFERTILIZED EGGS,  
AND THEN PROGRESSIVELY DECREASES TO BECOME UNDETECTABLE AT THE  
GASTRULA STAGE.  
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.  
CC -1- SIMILARITY: BELONGS TO THE MDM2 / MDM4 FAMILY.  
DR PDB: 1YCO; 19-NOV-97.  
DR InterPro: IPR003160; MDM2.  
DR InterPro: IPR001876; Znf-RanBP.  
DR InterPro: IPR001841; Znf\_ring.  
DR Pfam: PF02279; MDM2; 1.  
DR Pfam: PF00641; zf-RanBP; 1.  
DR SMART: SM00184; RING; 1.  
DR SMART: SM00547; ZNF\_RBZ; 1.  
DR PROSITE: PS01358; ZF-RANBP2\_1; 1.  
DR PROSITE: PS01358; ZF-RANBP2\_2; 1.  
DR PROSITE: PS00518; ZF\_RING\_1; FALSE\_NEG.  
DR PROSITE: PS00518; ZF\_RING\_2; 1.  
KW Nuclear protein; Ligase; Ubiquitin conjugation; Zinc; Zinc-finger;  
KW Metal-binding; 3D-structure.  
FT DOMAIN 15 104 REGION I.  
FT DOMAIN 173 179 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 230 230 REGION II.  
FT DOMAIN 216 292 ASP/GLU-RICH (ACIDIC).  
FT ZN\_FING 290 319 RANBP2-TYPE.  
FT ZN\_FING 420 461 RING-TYPE.  
SO SEQUENCE 473 AA; 53464 MW; 7DA668DEB3BEE01 CRC64;  
  
Query Match 55.6%; Score 35; DB 1; Length 473;  
Best Local Similarity 61.5%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 KTEIISVYNXVAE 13  
Db 160 ETEIISVDHPAE 172  
ID MALO\_STRPN STANDARD; PRT; 505 AA.  
AC P56273;

```

AC P29851;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE (Disproportionating enzyme) (D-enzyme).
GN MALQ OR MALM OR SP2107.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83129347; PubMed=6297760;
RA Lacks S.A., Dunn J.J., Greenberg B.;
RT "Identification of base mismatches recognized by the
RT heteroduplex-DNA-repair system of Streptococcus pneumoniae.";
RL Cell 31:327-336(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TIGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Knouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
CC -1- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan
CC to a new 4-position in an acceptor, which may be glucose or (1,4)-
CC alpha-D-glucan.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE DISPROPORTIONATING ENZYME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC DR EMBL: J01796; AAA26923.1; -
CC DR EMBL: AE007499; AAK76166.1; -
CC DR HSSP: 087172; 1CWW.
CC DR TIGR: SP2107; -
CC DR InterPro: IPR003385; 4A_gluca_ntrans.
CC DR Pfam: PF02446; 4A_gluca_ntrans; 1.
CC DR Transfaser: Glycosyltransferase; Carbohydrate metabolism;
CC Complete proteome.
CC KW SEQUENCE 505 AA; 58076 MW; D4529A000A6D01A5 CRC64;
CC SQ

```

```

Query Match 55.6%; Score 35; DB 1; Length 505;
Best Local Similarity 54.5%; Pred. No. 55;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 3 EEISEVNXVAE 13
   11:11:11
DB 332 EEIGELINIAE 342

```

```

RESULT 12
PAC_ARTVI STANDARD; PRT; 802 AA.
AC P31956;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE (Penicillin G amidohydrolase).
GN PAC OR PA.
OS Arthrobacter viscosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1673;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15294;
RX MEDLINE=94259306; PubMed=8200542;
RA Konstantinov N., Marjanovic N., Ujubi-Jankic G., Glisin V.;
RT "The penicillin amidase of Arthrobacter viscosus (ATCC 15294).";
RL Gene 143:79-83(1994).
CC -1- CATALYTIC ACTIVITY: Penicillin + H(2)O = a fatty acid anion + 6-
CC antihypocrellanate.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A BETA CHAIN
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Extracellular (potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S45 (SERINE PROTEASE) ALSO
CC KNOWN AS THE PENICILLIN ACYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC DR EMBL: L04471; AAA22077.1; -
CC DR HSSP: P06875; 1AQ.
CC DR MEROPS: S45.001; -
CC DR InterPro: IPR002692; Penicill_amidase.
CC DR Pfam: PF01804; Penicill_amidase; 1.
CC KM Hydrolyase; Antibiotic resistance; zymogen; signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 802 POTENTIAL.
CC FT CHAIN 25 234 PENICILLIN G ACYLASE ALPHA SUBUNIT.
CC FT PROPEP 235 265 SPACER PEPTIDE.
CC FT CHAIN 266 802 PENICILLIN G ACYLASE BETA SUBUNIT.
CC FT ACT_SITE 266 266 BY SIMILARITY.
CC SQ SEQUENCE 802 AA; 92113 MW; 96635E8C526C85D CRC64;
CC

```

```

Query Match 55.6%; Score 35; DB 1; Length 802;
Best Local Similarity 50.0%; Pred. No. 90;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 3 EEISEVNXVAE 14
   11:11:11
DB 548 EDINETNTYASF 559

```

```

RESULT 13
PAC_BACME STANDARD; PRT; 802 AA.
AC Q60136; Q9S463;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Penicillin G acylase precursor (EC 3.5.1.11) (Penicillin G amidase)
DE (Penicillin G amidohydrolase).
GN PAC OR PGA.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 14945;
RA Kang J.H., Kim S.J., Park Y.C., Hwang Y., Yoo O.J., Kim Y.C.;
RT "Nucleotide sequence of the penicillin G acylase gene from
RT Bacillus megaterium and characteristics of the enzyme.";

```

```

RL Misamunrag Holj 32:215-221(1994).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14945;
RX MEDLINE=95180705; PubMed=7875576;
RA Martin L.M., Prieto A.M., Cortes F., Garcia J.L.;
RT "Cloning and sequencing of the pac gene encoding the penicillin G
RT acylase of Bacillus megaterium ATCC 14945."
RL FEMS Microbiol. Lett. 125:287-292(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CA4098;
RA Yang S., Huang X.D., Huang Y.H., Li S.Y., Yuan Z.Y.;
RT "Bacillus megaterium CA4098 penicillin G amidase structural gene
RT sequence."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: Penicillin + H(2)O -> a fatty acid anion + 6-
CC amphenicolanate.
CC -1 SUBUNIT: HETERODIMER OF AN ALPHA CHAIN AND A BETA CHAIN
CC (BY SIMILARITY).
CC -1 SUBCELLULAR LOCATION: Extracellular (Potential).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S45 (SERINE PROTEASE) ALSO
CC KNOWN AS THE PENICILLIN ACYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U07682; AAB41343.1; -
CC EMBL: 237542; CA85774.1; -
CC DR EMBL: AF161313; AAD45609.1; -
CC DR HSSP: P06875; IAO.
CC DR MEROPS: S45.001; -
CC DR InterPro: IPR002692; Penicill_amidase.
CC Pfam: PF01804; Penicill_amidase; 1.
CC K1M Hydrolyase; Antibiotic resistance; zymogen; signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 802
CC FT CHAIN 25 234
CC FT PROPEP 235 265
CC FT CHAIN 266 802
CC FT ACT_SITE 266 266
CC FT VARIANT 3 3
CC FT VARIANT 224 224
CC FT VARIANT 232 232
CC FT VARIANT 254 254
CC FT VARIANT 349 349
CC FT VARIANT 470 470
CC FT VARIANT 524 524
CC FT VARIANT 569 569
CC FT VARIANT 586 586
CC FT VARIANT 657 657
CC FT VARIANT 740 740
CC FT VARIANT 789 791
CC FT SEQUENCE 802 AA; 91987 MW; 877CA0564E50DFD CRC64;
CC
CC Query Match 55.6%; Score 35; DB 1; Length 802;
CC Best Local Similarity 50.0%; Pred. No. 90;
CC Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Colanic acid capsular biosynthesis activation protein A.
GN RCSA.
OS Klebsiella aerogenes.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCB1_TaxID=28451.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88009858; PubMed=3309150;
RA Allen P., Hart C.A., Saunders J.R.;
RT "Isolation from Klebsiella and characterization of two rcs genes that
RT activate colanic acid capsular biosynthesis in Escherichia coli."
RL J. Gen. Microbiol. 133:331-340(1987).
CC -1 FUNCTION: POSITIVE REGULATOR OF CAPSULAR POLYSACCHARIDE
CC SYNTHESIS. RCSA AND RCSB FORM A COMPLEX TO PROMOTE TRANSCRIPTION
CC OF THE GENES FOR CAPSULE SYNTHESIS.
CC -1 SIMILARITY: BELONGS TO THE LUXR/DIRA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15748; AAA25142.1; -
CC DR PIR: A47615; A47615.
CC DR InterPro: IPR000792; HTH_LuxR.
CC DR Pfam: PF00196; Gete; 1.
CC DR PRINTS: PR00038; HTHLUXR.
CC DR SMART: SM00421; HTH_LUXR; 1.
CC DR PROSITE: PS00622; HTH_LUXR_FAMILY; 1.
CC K1M Transcription regulation; DNA-binding; Activator.
CC FT DNA_BIND 155 174
CC FT SEQUENCE 207 AA; 23366 MW; 24E2253B9DEE162 CRC64;
CC
CC Query Match 54.0%; Score 34; DB 1; Length 207;
CC Best Local Similarity 53.8%; Pred. No. 33;
CC Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: TO YEAST YOR062C.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U01878; AAA79998.1; -
DR EMBL; Z28300; CAA82154.1; -
DR PIR; S38152; S38152.
DR SGD; S0001783; YKR075C.
KW Hypothetical protein.
FT DOMAIN 254 270
SQ SEQUENCE 307 AA; 36113 MW; 0DC6D0241DADE1C5 CRC64;
-----
Query Match 54.0%; Score 34; DB 1; Length 307;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 TPEISEVNVAE 13
DB 157 TEEVDEINASTE 168

```

Search completed: October 30, 2002, 12:27:47  
Job time : 4.20393 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 4.36855 Seconds

(Without alignments)  
554,401 Million cell updates/sec

Title: US-09-724-571-72

Perfect score: 63  
Sequence: 1 KTEISEVKNVAEF 14

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_fodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriaph:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	76.2	79	11 035463	035463 cricetus
2	48	76.2	82	4 P78438	P78438 homo sapien
3	48	76.2	82	4 Q16014	Q16014 homo sapien
4	48	76.2	82	4 Q16019	Q16019 homo sapien
5	48	76.2	82	4 Q16020	Q16020 homo sapien
6	48	76.2	607	11 099K32	099K32 mus musculu
7	48	76.2	695	6 Q95K87	Q95K87 macaca fasc
8	48	76.2	695	11 P97487	P97487 mus musculu
9	48	76.2	695	11 Q60496	Q60496 cavia sp. p
10	48	76.2	770	6 Q9TU10	Q9TU10 sus scrofa
11	47	74.6	534	13 Q93296	Q93296 gallus gall
12	47	74.6	569	13 Q9PVL1	Q9PVL1 gallus gall
13	47	74.6	695	13 Q9DCJ8	Q9DCJ8 gallus gall
14	47	74.6	751	13 Q9DCJ7	Q9DCJ7 gallus gall
15	42	66.7	693	13 Q98SG0	Q98SG0 xenopus lae
16	42	66.7	747	13 Q91963	Q91963 xenopus. ap

17	39	61.9	699	13 057394	057394 narke japon
18	38	60.3	316	5 Q9XVK7	Q9XVK7 caenorhabdi
19	38	60.3	335	5 Q96847	Q96847 dictyosteli
20	38	60.3	335	5 Q23896	Q23896 dictyosteli
21	38	60.3	1334	5 Q9V6K1	Q9V6K1 drosophila
22	38	60.3	1354	5 Q9V6K1	Q9V6K1 drosophila
23	38	60.3	1803	5 Q9VU16	Q9VU16 drosophila
24	37.5	59.5	213	9 Q9XJ88	Q9XJ88 bacterioph
25	37	58.7	561	5 Q950M4	Q950M4 caenorhabdi
26	37	58.7	564	9 Q9AZ64	Q9AZ64 lactococcus
27	37	58.7	568	4 Q96LQ3	Q96LQ3 homo sapien
28	37	58.7	946	10 Q9FE91	Q9FE91 arabisdopsi
29	36.5	57.9	587	10 Q9LYU1	Q9LYU1 arabisdopsi
30	36	57.1	180	16 Q92P99	Q92P99 rhizobium m
31	36	57.1	190	12 Q9EMT2	Q9EMT2 amsaeta moo
32	36	57.1	354	5 Q9VER2	Q9VER2 drosophila
33	36	57.1	415	5 Q44228	Q44228 halocynthia
34	36	57.1	470	8 Q9MSU0	Q9MSU0 juniperus c
35	36	57.1	502	8 Q9MST8	Q9MST8 juniperus p
36	36	57.1	512	10 Q9C8H8	Q9C8H8 arabisdopsi
37	36	57.1	513	10 Q38821	Q38821 arabisdopsi
38	36	57.1	570	10 Q9PCT6	Q9PCT6 caenorhabdi
39	36	57.1	650	5 Q17866	Q17866 caenorhabdi
40	36	57.1	695	13 Q98SE9	Q98SE9 xenopus lae
41	36	57.1	778	5 Q9U9K6	Q9U9K6 caenorhabdi
42	36	57.1	837	5 Q25751	Q25751 plasmodium
43	36	57.1	864	5 Q93336	Q93336 caenorhabdi
44	36	57.1	867	10 Q81747	Q81747 arabisdopsi
45	36	57.1	1117	5 Q9U9K7	Q9U9K7 caenorhabdi

## ALIGNMENTS

## RESULT 1

ID 035463 PRELIMINARY; PRT; 79 AA.  
AC 035463;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ALHEIMER'S AMYLOID BETA PROTEIN (FRAGMENT).  
OS BETA APP.  
GN Cricetus griseus (Chinese hamster).  
OC Cricetidae; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
CC Cricetus.  
OX NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sambamurti K., Plinix I., Gandhi S.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF030413; AAB86608.1; -.  
DR HSSP: P05067; IBA4.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 79 AA: 8538 MW: 37F2C6C3BFF3F597 CRC64;

Query Match 76.2%; Score 48; DB 11; Length 79;  
Best Local Similarity 78.6%; Pred. NO. 0.19;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVKNVAEF 14  
Db 11 KTEISEVKNVAEF 24

RESULT 2  
ID P78438 PRELIMINARY; PRT; 82 AA.  
AC P78438;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

```

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE AMYLOID PROTEIN (BETA-AMYLOID PROTEIN) (FRAGMENT).
GN APP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89392030; PubMed=2675837;
RA Johnstone E.M., Chaney M.O., Moore R.E., Ward K.E., Norris F.H.,
RA Little S.P., Alzheimer's disease amyloid peptide is encoded by two exons and shows
RT "Alzheimer's disease amyloid peptide is encoded by two exons and shows
RT similarity to soybean trypsin inhibitor.";
RL Biochem. Biophys. Res. Commun. 163:1248-1255(1989).
RN [2]
RP SEQUENCE OF 19-48 FROM N.A.
RX MEDLINE=87120329; PubMed=2949367;
RA Tanzil R.E., Gusella J.F., Watkins P.C., Bruns G.A., George-Hyslop P.,
RA Van Keuren M.L., Patterson D., Pagan S., Kurit D.M., Neve R.L.;
RT "Amyloid beta protein gene: cDNA, mRNA distribution, and genetic
RT linkage near the Alzheimer locus.";
RL Science 235:880-884(1987).
RN [3]
RP SEQUENCE OF 32-63 FROM N.A.
RX MEDLINE=93035397; PubMed=1415269;
RA Kamino K., Orr H.T., Payant H., Wjisman E.M., Alonso M.E., Pulst S.M.,
RA Anderson L., O'dahl S., Nemens E., White J.A.;
RT "Linkage and mutational analysis of familial Alzheimer disease
RT kindreds for the APP gene region.";
RL Am. J. Hum. Genet. 51:998-1014(1992).
DR EMBL; M29270; AAA51768.1; JOINED.
DR EMBL; M29269; AAA51768.1; JOINED.
DR EMBL; M15532; AAA51564.1; -.
DR EMBL; S45136; AAB23646.1; -.
DR HSSP; P05067; 1BA4.
FT NON_TER
SQ SEQUENCE 82 AA; 8994 MW; 8DA9E42B813A070E CRC64;

Query Match 76.2%; Score 48; DB 4; Length 82;
Best Local Similarity 78.6%; Pred. No. 0.19;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14
DB 7 KTEISEVKNDAEF 20
ID 016014 PRELIMINARY; PRT; 82 AA.
AC 016014;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BETA-AMYLOID PEPTIDE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B., Rosenzwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease
RT mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S60721; AAB26263.2; -.
DR HSSP; P05067; 1BA4.
FT NON_TER
SQ SEQUENCE 82 AA; 8972 MW; F534AA5B3EA9230A CRC64;

```

```

Query Match 76.2%; Score 48; DB 4; Length 82;
Best Local Similarity 78.6%; Pred. No. 0.19;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14
DB 8 KTEISEVKNDAEF 21
ID 016019 PRELIMINARY; PRT; 82 AA.
AC 016019;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BETA-AMYLOID PEPTIDE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B., Rosenzwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease
RT mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61380; AAB26264.2; -.
DR HSSP; P05067; 1BA4.
FT NON_TER
SQ SEQUENCE 82 AA; 8938 MW; F534AA50E579230A CRC64;

Query Match 76.2%; Score 48; DB 4; Length 82;
Best Local Similarity 78.6%; Pred. No. 0.19;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14
DB 8 KTEISEVKNDAEF 21
ID 016020 PRELIMINARY; PRT; 82 AA.
AC 016020;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BETA-AMYLOID PEPTIDE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93236601; PubMed=8476439;
RA Denman R.B., Rosenzwaig R., Miller D.L.;
RT "A system for studying the effect(s) of familial Alzheimer disease
RT mutations on the processing of the beta-amyloid peptide precursor.";
RL Biochem. Biophys. Res. Commun. 192:96-103(1993).
DR EMBL; S61383; AAB26265.2; -.
DR HSSP; P05067; 1BA4.
FT NON_TER
SQ SEQUENCE 82 AA; 8882 MW; F534AA5AE5D9230A CRC64;

Query Match 76.2%; Score 48; DB 4; Length 82;
Best Local Similarity 78.6%; Pred. No. 0.19;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Oy 1 KTEISEVNXVAEF 14  
 ||||||| 111  
 Db 8 KTEISEVKMDAEF 21

## RESULT 6

099K32 ID 099K32 PRELIMINARY; PRT; 607 AA.

AC 099K32, 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL 68.4 KDA PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS  
 RC TISSUE.  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC005490; AAH05490.1; -  
 DR HSSP; P05067; 1AAP.  
 DR InterPro: IPR001868; A4\_APP.  
 DR Pfam: PF00014; Kunitz\_BPTI.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR SMART; SM00131; KU; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS0279; BPTI\_KUNITZ\_2; 1.  
 KW Hypothetical protein; Serine protease inhibitor.  
 FT NON-TER 1  
 SQ SEQUENCE 607 AA; 68391 MW; BF02214CBA7D172 CRC64;

## Query Match

Best Local Similarity 76.2%; Score 48; DB 11; Length 607;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KTEISEVNXVAEF 14  
 ||||||| 111  
 Db 499 KTEISEVKMDAEF 512

## RESULT 7

095KN7 ID 095KN7 PRELIMINARY; PRT; 695 AA.

AC 095KN7, 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE AMYLOID B-PROTEIN PRECURSOR.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE-CEREBELLUM;  
 RX MEDLINE=91273117; PubMed=1905108;  
 RA Podlisky M.B., Tolian D.R., Selkoe D.J.;  
 RT "Homology of the amyloid beta protein precursor in monkey and human  
 disease";  
 RT supports a primate model for beta amyloidosis in Alzheimer's  
 disease.  
 RL Am. J. Pathol. 138:1423-1435(1991).  
 DR EMBL; M58727; AAA36829.1; -  
 FT SIGNAL 1 17 POTENTIAL.

FT CHAIN 597 636 POTENTIAL.  
 SQ SEQUENCE 695 AA; 78663 MW; 4F6DA0139F69D56 CRC64;

Query Match  
 Best Local Similarity 76.2%; Score 48; DB 6; Length 695;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KTEISEVNXVAEF 14  
 ||||||| 111  
 Db 587 KTEISEVKMDAEF 600

## RESULT 8

P97487 ID P97487 PRELIMINARY; PRT; 695 AA.

AC P97487, 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HIPPOCAMPAL AMYLOID PROTEIN.  
 GN APP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=SAMP8; TISSUE=HIPPOCAMPUS;  
 RA Flood J.F., Kumar V.B., Sasser T., Word I., Morley J.E.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RC SEQUENCE OF 581-662 FROM N.A.  
 RC STRAIN=129SV;  
 RA Wragg M.A., Busfield F., Duff K., Korenblatt K., Capecchi M.,  
 RA Loring J.F., Goate A.M.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U84012; AAB41502.1; -  
 DR EMBL; U82624; AAB40919.1; -  
 DR HSSP; P05067; 1MWP.  
 DR MGD; MGI:88059; APP.  
 DR InterPro: IPR001868; A4\_APP.  
 DR Pfam; PF02177; A4\_EXTRA; 1.  
 DR PRINTS; PR00203; AMYLOIDA4.  
 DR SMART; SM00006; A4\_EXTRA; 1.  
 DR PROSITE; PS00319; A4\_EXTRA; 1.  
 DR PROSITE; PS00320; A4\_INTRA; 1.  
 SQ SEQUENCE 695 AA; 78414 MW; 9A5FB2BD261236E CRC64;

## Query Match

Best Local Similarity 76.2%; Score 48; DB 11; Length 695;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KTEISEVNXVAEF 14  
 ||||||| 111  
 Db 587 KTEISEVKMDAEF 600

## RESULT 9

060496 ID 060496 PRELIMINARY; PRT; 695 AA.

AC 060496, 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PUTATIVE AMYLOID PRECURSOR PROTEIN.  
 OS Cavia sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.  
 OX NCBI\_TaxID=10143;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=97236426; PubMed=9116031;

RA Beck M., Mueller D., Bigl V.  
 RT "Amyloid precursor protein in Guinea pigs - complete cDNA sequence and  
 alternative splicing."  
 RL Biochim. Biophys. Acta 1351:17-21(1997).  
 DR EMBL: X97631; CA66230.1; -.  
 DR HSSP: P05067; 1BA4.  
 DR InterPro: IPR001868; A4\_APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 SO SEQUENCE 695 AA; 78701 MW; 5196A0C4017F16AB CRC64;

Query Match 76.28; Score 48; DB 11; Length 695;  
 Best Local Similarity 78.68; Pred. No. 1.7;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 587 KTEISEVKMDAEF 600

RESULT 10  
 O9TUI0 PRELIMINARY; PRT; 770 AA.  
 AC O9TUI0:  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE AMYLOID PRECURSOR PROTEIN.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euteria; Cetartiodactyla; Suina; Suidae; Sus.  
 NCBI\_TaxID=9823;  
 RN NCBI\_TaxID=9823;  
 RP SEQUENCE FROM N.A.  
 RA Kimura A., Takahashi T.  
 RT "Amyloid Precursor Protein 770."  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB032550; BAA84580.1; -.  
 DR HSSP: P05067; 1AAP.  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR Pfam: PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR PRINTS: PR00759; BASICTPASE.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR SMART: SM00131; KU; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_2; 1.  
 DR PROSITE: PS00279; BPTI\_KUNITZ\_2; 1.  
 KM Serine protease inhibitor.  
 SO SEQUENCE 770 AA; 86961 MW; 5F7A1DCB2BC583E CRC64;

Query Match 76.28; Score 48; DB 6; Length 770;  
 Best Local Similarity 78.68; Pred. No. 1.8;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 662 KTEISEVKMDAEF 675

RESULT 11  
 O93296 PRELIMINARY; PRT; 534 AA.  
 AC O93296:  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE AMYLOID PROTEIN (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN NCBI\_TaxID=9031;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE:98337885; PubMed:9671674;  
 RA Barnes N.Y., Li L., Yoshikawa K., Schwartz L.M., Oppenheim R.W.,  
 RA Milligan C.E.;  
 RT "Increased production of amyloid precursor protein provides a  
 substrate for caspase-3 in dying motoneurons."  
 RL J. Neurosci. 18:5869-5880(1998).  
 DR EMBL: AF042098; AAC25052.1; -.  
 DR HSSP: P05067; 1BA4.  
 DR InterPro: IPR001868; A4\_APP.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 FT NON TER 1  
 SO SEQUENCE 534 AA; 60597 MW; FB53ECC2E66D4C92 CRC64;

Query Match 74.68; Score 47; DB 13; Length 534;  
 Best Local Similarity 71.48; Pred. No. 2;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 426 KTEISEVKMDAEF 439

RESULT 12  
 O9PVL1 PRELIMINARY; PRT; 569 AA.  
 AC O9PVL1:  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE AMYLOID PROTEIN (FRAGMENT).  
 GN App.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN NCBI\_TaxID=9031;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Coulson E.J., Paliaga K., Beyreuther K., Masters C.L.;  
 RT "What the evolution of the amyloid protein precursor supergene family  
 tells us about its function."  
 RL Neurochem. Int. 0:0-0(2000).  
 DR EMBL: AF030341; AAF12698.1; -.  
 DR HSSP: P05067; 1BA4.  
 DR InterPro: IPR001868; A4\_APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR PRINTS: PR00203; AMYLOIDA4.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 FT NON TER 1  
 SO SEQUENCE 569 AA; 64753 MW; 0AB8BB851863A19D CRC64;

Query Match 74.68; Score 47; DB 13; Length 569;  
 Best Local Similarity 71.48; Pred. No. 2.1;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KTEISEVNXVAEF 14  
 |||||  
 DB 462 KTEISEVKMDAEF 475

## RESULT 13

09DCJ8 ID Q9DCJ8 PRELIMINARY: PRT: 695 AA.  
 AC Q9DCJ8;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-AMYLOID PRECURSOR PROTEIN 695 ISOFORM.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sarasa M., Rodolose A., Sorribas V.;  
 RT "Cloning of full-length chicken beta-amyloid precursor protein  
 RT isoforms."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF289218; AAG00593.1; -;  
 DR HSSP: P05067; 1BA4.  
 DR InterPro: IPR001868; A4\_APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR PRINTS: PRO0203; AMYLOIDA4.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 SQ SEQUENCE 695 AA; 78565 MW; F201ED02AEC86D95 CRC64;

## Query Match

Best Local Similarity 74.6%; Score 47; DB 13; Length 695;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
 DB 587 KTEEVEVKMDAEF 600

## RESULT 14

09DCJ7 ID Q9DCJ7 PRELIMINARY: PRT: 751 AA.  
 AC Q9DCJ7;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-AMYLOID PRECURSOR PROTEIN 751 ISOFORM.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sarasa M., Rodolose A., Sorribas V.;  
 RT "Cloning of full-length chicken beta-amyloid precursor protein  
 RT isoforms."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF289219; AAG00594.1; -;  
 DR HSSP: P05067; 1BA4.  
 DR InterPro: IPR001868; A4\_APP.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR PRINTS: PRO0203; AMYLOIDA4.  
 DR PRINTS: PRO0759; BASICTPASE.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR SMART: SM00131; KU; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS0279; BPTI\_KUNITZ\_2; 1.  
 KW Serine protease inhibitor.  
 SQ SEQUENCE 751 AA; 84705 MW; E78E9413A8033D84 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 751;  
 Best Local Similarity 71.4%; Pred. No. 2.8;

Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
 DB 643 KTEEVEVKMDAEF 656

## RESULT 15

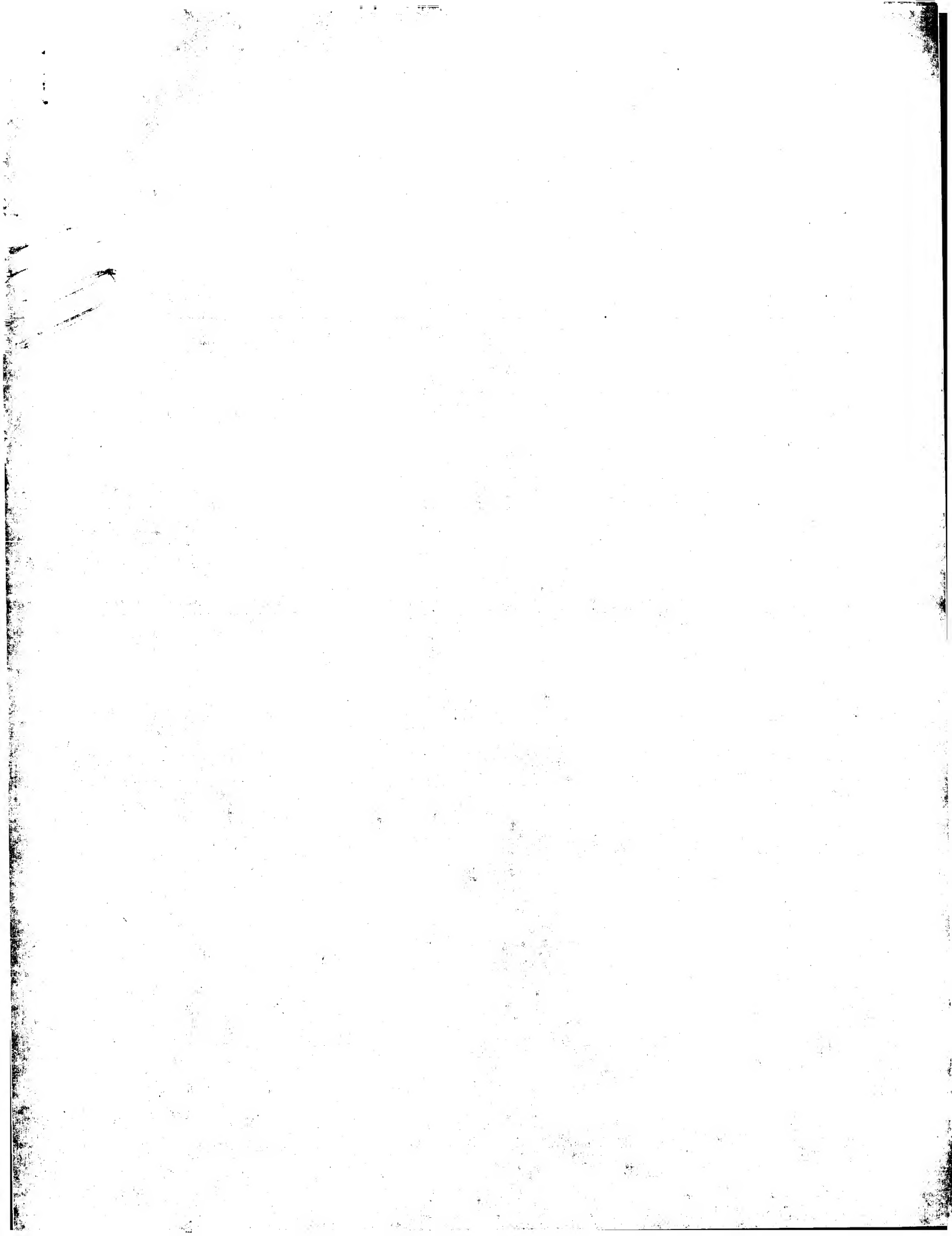
098SGO ID Q98SGO PRELIMINARY: PRT: 693 AA.  
 AC Q98SGO;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE BETA-AMYLOID PRECURSOR PROTEIN A.  
 GN APP.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8395;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Van den Hark W.H.;  
 RL Thesis (2001), Department of Biological Sciences,  
 RL University of Nijmegen, Nijmegen, Netherlands.  
 DR EMBL: AJ28150; CAC37193.1; -;  
 DR HSSP: P05067; 1H23.  
 DR InterPro: IPR001868; A4\_APP.  
 DR Pfam: PF02177; A4\_EXTRA; 1.  
 DR PRINTS: PRO0203; AMYLOIDA4.  
 DR SMART: SM00006; A4\_EXTRA; 1.  
 DR PROSITE: PS00319; A4\_EXTRA; 1.  
 DR PROSITE: PS00320; A4\_INTRA; 1.  
 KW Signal.  
 FT SIGNAL.  
 SQ SEQUENCE 693 AA; 78568 MW; CAF1DF655C1AB653 CRC64;

## Query Match

Best Local Similarity 66.7%; Score 42; DB 13; Length 693;  
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KTEISEVNXVAEF 14  
 DB 585 KTEEVEVKMDSEY 598

Search completed: October 30, 2002, 12:30:01  
 Job time : 6.36855 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 ; Search time 0.997543 Seconds  
(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-78

Perfect score: 29

Sequence: 1 VMXAEF 7

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_patents\_AA.\*  
2: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	82.8	972	3	US-08-335-844A-24
2	23	79.3	54	2	US-08-456-647B-18
3	23	79.3	54	2	US-08-456-647B-46
4	23	79.3	54	2	US-08-456-647B-47
5	23	79.3	54	2	US-08-237-401A-18
6	23	79.3	54	2	US-08-237-401A-46
7	23	79.3	54	2	US-08-237-401A-47
8	23	79.3	299	2	US-08-701-191A-13
9	23	79.3	300	2	US-08-701-191A-31
10	23	79.3	309	2	US-08-701-191A-9
11	23	79.3	310	2	US-08-701-191A-1
12	23	79.3	310	2	US-08-701-191A-6
13	23	79.3	310	2	US-08-701-191A-7
14	23	79.3	310	2	US-08-701-191A-8
15	23	79.3	310	4	US-09-390-326-7
16	23	79.3	313	1	US-08-278-089A-17
17	23	79.3	313	1	US-08-838-957A-16
18	23	79.3	315	2	US-08-701-191A-2
19	23	79.3	318	2	US-08-701-191A-10
20	23	79.3	351	2	US-08-701-191A-3
21	23	79.3	378	1	US-08-070-165F-8
22	23	79.3	378	2	US-08-885-418-8
23	23	79.3	388	1	US-08-070-165F-4
24	23	79.3	388	2	US-08-885-418-4
25	23	79.3	612	2	US-08-673-789-11
26	23	79.3	652	1	US-08-471-570-10
27	23	79.3	729	1	US-07-640-029-3

28	23	79.3	729	1	US-08-070-165F-6	Sequence 6, Appl1
29	23	79.3	729	2	US-08-885-418-6	Sequence 6, Appl1
30	23	79.3	731	1	US-07-921-807B-5	Sequence 5, Appl1
31	23	79.3	731	1	US-08-070-165F-10	Sequence 10, Appl1
32	23	79.3	731	1	US-08-441-944A-5	Sequence 5, Appl1
33	23	79.3	731	2	US-08-885-418-10	Sequence 10, Appl1
34	23	79.3	731	4	US-08-439-992A-3	Sequence 3, Appl1
35	23	79.3	733	1	US-07-640-029-4	Sequence 4, Appl1
36	23	79.3	733	1	US-07-921-807B-6	Sequence 6, Appl1
37	23	79.3	733	1	US-08-441-944A-4	Sequence 4, Appl1
38	23	79.3	733	4	US-08-439-992A-4	Sequence 4, Appl1
39	23	79.3	769	1	US-08-471-570-8	Sequence 8, Appl1
40	23	79.3	801	4	US-09-383-630-6	Sequence 6, Appl1
41	23	79.3	802	4	US-09-173-151A-33	Sequence 3, Appl1
42	23	79.3	806	4	US-09-383-630-3	Sequence 3, Appl1
43	23	79.3	816	1	US-07-640-029-1	Sequence 1, Appl1
44	23	79.3	817	1	US-07-640-029-2	Sequence 2, Appl1
45	23	79.3	820	1	US-07-921-807B-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-335-844A-24  
Sequence 24, Application US/08335844A  
Patent No. 6066503  
GENERAL INFORMATION:  
APPLICANT: GRAHAM, MARGARET  
APPLICANT: SMITH, TREVOR STANLEY  
APPLICANT: MUNN, EDWARD ALBERT  
APPLICANT: KNOX, DAVID PATRICK  
APPLICANT: OLIVER, JOANNA JANE  
APPLICANT: NEWTON, SUSAN ELIZABETH  
TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING  
TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE  
TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz  
STREET: Suite 701-E, 555 Thirteenth St., N.W  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/335,844A  
FILING DATE: 09-JAN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB PCT/GB93/00943  
FILING DATE: 06-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9209936  
FILING DATE: 08-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WALKER, Barbara W.  
REGISTRATION NUMBER: 35,400  
REFERENCE/DOCKET NUMBER: 1181-223A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)783-6040  
TELEFAX: (202)783-6031  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 972 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-335-844A-24

Query Match 82.8%; Score 24; DB 3; Length 972;  
Best Local Similarity 71.4%; Pred. No. 3,4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7

Db 542 VISVAEF 548

RESULT 2

US-08-456-647B-18

; Sequence 18, Application US/08456647B

; Patent No. 5811516

; GENERAL INFORMATION:

; APPLICANT: Lemke Ph.D. et al., Greg E.

; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/456,647B

; FILING DATE: 02-JUN-1995

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/237,401

; FILING DATE: 02-MAY-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/884,486

; FILING DATE: 15-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Wetherell Ph.D., John R.

; REGISTRATION NUMBER: 31,678

; REFERENCE/DOCKET NUMBER: 07251/007002

; TELEPHONE: (619) 678-5070

; TELEFAX: (619) 678-5099

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-456-647B-18

Query Match 79.3%; Score 23; DB 2; Length 54;

Best Local Similarity 57.1%; Pred. No. 28;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7

Db 9 VMKIADF 15

RESULT 3

US-08-456-647B-46

; Sequence 46, Application US/08456647B

; Patent No. 5811516

; GENERAL INFORMATION:

; APPLICANT: Lemke Ph.D. et al., Greg E.

; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES

; NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish &amp; Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/456,647B

FILING DATE: 02-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/237,401

FILING DATE: 02-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/884,486

FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wetherell Ph.D., John R.

REGISTRATION NUMBER: 31,678

REFERENCE/DOCKET NUMBER: 07251/007002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 54 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-456-647B-46

Query Match 79.3%; Score 23; DB 2; Length 54;

Best Local Similarity 57.1%; Pred. No. 28;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7

Db 9 VMKIADF 15

RESULT 4

US-08-456-647B-47

; Sequence 47, Application US/08456647B

; Patent No. 5811516

; GENERAL INFORMATION:

; APPLICANT: Lemke Ph.D. et al., Greg E.

; TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/456,647B

; FILING DATE: 02-JUN-1995

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/237,401



FILED DATE: 02-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetterell Ph.D., John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: 07251/007002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-456-647B-47

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 28;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
| | : | : |  
DB 9 VKRIADF 15

RESULT 5  
US-08-237-401A-18  
Sequence 18, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237,401A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-18

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 28;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VMXVAEF 7  
| | : | : |  
DB 9 VKRIADF 15

RESULT 6  
US-08-237-401A-46  
Sequence 46, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237,401A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-46

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 28;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
| | : | : |  
DB 9 VKRIADF 15

RESULT 7  
US-08-237-401A-47  
Sequence 47, Application US/08237401A  
Patent No. 5837448  
GENERAL INFORMATION:  
APPLICANT: Lemke Ph.D. et al., Greg E.  
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US

ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/237,401A  
FILING DATE: 02-May-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/884,486  
FILING DATE: 15-May-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07251/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-237-401A-47

Query Match 79.3%; Score 23; DB 2; Length 54;  
Best Local Similarity 57.1%; Pred. No. 28;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXYAEF 7  
||:|:|  
DB 9 VMKIADF 15

RESULT 8  
US-08-701-191A-13  
Sequence 31, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Steven R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 299 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-31

Query Match 79.3%; Score 23; DB 2; Length 299;  
Best Local Similarity 57.1%; Pred. No. 1.7e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXYAEF 7  
||:|:|  
DB 172 VMKIADF 178

RESULT 9  
US-08-701-191A-31  
Sequence 31, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Steven R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 955-0440  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 300 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-31

Query Match 79.3%; Score 23; DB 2; Length 300;  
Best Local Similarity 57.1%; Pred. No. 1.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAF 7  
Db 172 VMKIDF 178

## RESULT 10

US-08-701-191A-9  
Sequence 9, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 309 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-9

Query Match 79.3%; Score 23; DB 2; Length 309;  
Best Local Similarity 57.1%; Pred. No. 1.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAF 7  
Db 181 VMKIDF 187

RESULT 11  
US-08-701-191A-1  
Sequence 1, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-701-191A-1

Query Match 79.3%; Score 23; DB 2; Length 310;  
Best Local Similarity 57.1%; Pred. No. 1.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAF 7  
Db 181 VMKIDF 187

RESULT 12  
US-08-701-191A-6  
Sequence 6, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-6

Query Match  
Best Local Similarity 57.1%; Score 23; DB 2; Length 310;  
Pred. No. 1.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:|:|  
Db 181 VMKIADF 187

RESULT 13  
US-08-701-191A-7  
Sequence 7, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
APPLICANT: and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:

LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-7

Query Match  
Best Local Similarity 57.1%; Score 23; DB 2; Length 310;  
Pred. No. 1.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:|:|  
Db 181 VMKIADF 187

RESULT 14  
US-08-701-191A-8  
Sequence 8, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
APPLICANT: and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FASTSEQ for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-701-191A-8

Query Match  
Best Local Similarity 57.1%; Score 23; DB 2; Length 310;  
Pred. No. 1.8e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:|:|  
Db 181 VMKIADF 187

```

RESULT 15
US-09-390-326-7
; Sequence 7, Application US/09390326
; Patent No. 6316603
; GENERAL INFORMATION:
; APPLICANT: MCTIGUE, MICHELE A.
; APPLICANT: WICKERSHAM, JOHN A.
; APPLICANT: PINKO, CHRIS
; APPLICANT: SHOWALTER, RICHARD
; APPLICANT: PARAST, CAMRAN V.
; APPLICANT: TEMPCZYK-RUSSELL, ANNA
; APPLICANT: GEHRING, MICHAEL R.
; APPLICANT: MROCKOWSKI, BARBARA
; APPLICANT: KAN, CHEN-CHEN
; APPLICANT: VILLAFRANCA, J. ERNEST
; APPLICANT: APPELT, KRYSZTOF
; TITLE OF INVENTION: MODIFICATIONS OF THE VEGF RECEPTOR-2 PROTEIN AND
; FILE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 0125-0016US
; CURRENT APPLICATION NUMBER: US/09/390,326
; CURRENT FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-390-326-7

Query Match          79.3% Score 23; DB 4; Length 310;
Best Local Similarity 57.1%; Pred. NO. 1.8e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY      1 VMXVAEF 7
      ||:|:|
Db      181 VMKIADF 187

```

Search completed: October 30, 2002, 12:32:33  
 Job time : 1.99754 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:20:50 ; Search time 2.80344 Seconds  
(without alignments)  
277.344 Million cell updates/sec

Title: US-09-724-571-78

Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_032802:\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	93.1	7	21	AA07871
2	27	93.1	8	21	AA07872
3	26	89.7	821	20	AAV34479
4	26	89.7	869	20	AAV34354
5	24	82.8	75	22	AAU56778
6	24	82.8	132	22	AA41984
7	24	82.8	132	22	AA41985
8	24	82.8	132	22	AA41986
9	24	82.8	225	19	AAV85820
10	24	82.8	236	22	AAU50802
11	24	82.8	254	20	AAV36930

12	24	82.8	286	22	ABG18334	Novel human diago
13	24	82.8	335	22	AAU37753	Streptococcus pneu
14	24	82.8	361	21	AA020945	Arabidopsis thalia
15	24	82.8	361	21	AA024458	Arabidopsis thalia
16	24	82.8	361	21	AA045883	Arabidopsis thalia
17	24	82.8	373	22	AA040200	Human polypeptide
18	24	82.8	378	21	AA020944	Arabidopsis thalia
19	24	82.8	378	21	AA024457	Arabidopsis thalia
20	24	82.8	378	21	AA045882	Arabidopsis thalia
21	24	82.8	384	21	AA071036	Human membrane tra
22	24	82.8	384	22	AA040198	Human polypeptide
23	24	82.8	384	22	AA040199	Human polypeptide
24	24	82.8	414	21	AA020943	Arabidopsis thalia
25	24	82.8	414	21	AA024456	Arabidopsis thalia
26	24	82.8	414	21	AA045881	Arabidopsis thalia
27	24	82.8	584	22	AA06296	Putative P. abyss
28	24	82.8	817	22	AA070027	Drosophila melanog
29	24	82.8	962	15	AA08701	Aminopeptidase H1
30	24	82.8	972	14	AA051280	Helminth aminopept
31	23	79.3	36	18	AA027782	UDP-N-acetylglucos
32	23	79.3	54	19	AA079151	Receptor protein t
33	23	79.3	54	20	AA081408	Receptor protein t
34	23	79.3	84	22	AA041413	Propionibacterium
35	23	79.3	90	22	AA087752	Human T2R22 amino
36	23	79.3	143	21	AA010816	Archaeoglobus fulg
37	23	79.3	150	22	AA033403	Enterococcus faeca
38	23	79.3	189	22	AA052817	Escherichia coli p
39	23	79.3	237	22	AA090704	C glutamicum prote
40	23	79.3	246	22	AA058040	Drosophila melanog
41	23	79.3	249	22	AA013127	Sheep SLS FGFR3 no
42	23	79.3	249	22	AA013136	Sheep SLS FGFR3 mu
43	23	79.3	280	23	AA030751	C glutamicum prote
44	23	79.3	287	20	AA030631	Amino acid sequenc
45	23	79.3	299	22	AA063276	Drosophila melanog

ALIGNMENTS

RESULT 1	AA07871	standard; peptide: 7 AA.
XX	AA07871:	
AC	14-NOV-2000	(first entry)
XX		
DT		
XX		
DE		A beta-secretase inhibitor peptide.
XX		
KW		Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
KW		amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
KW		inhibitor.
XX		
OS		Synthetic.
XX		
PH	Key	Location/Qualifiers
FT	Modified-site	3 /note="hydroxyethylene"
XX		
PN	W0200047618-A2.	
XX		
PD	17-AUG-2000.	
XX		
PF	10-FEB-2000; 2000WO-US03819.	
XX		
PR	10-FEB-1999; 99US-0119571.	
PR	15-JUN-1999; 99US-0139172.	
XX		
PA	(ELAN-) ELAN PHARM INC.	
XX		
XX	Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;	
PI	Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;	
XX		





CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 XX  
 SQ Sequence 821 AA;  
 Query Match 89.7%; Score 26; DB 20; Length 821;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 VMXVAEF 7  
 Db 565 VMPVAEF 571  
 RESULT 4  
 AA134354  
 ID AA134354 standard; Protein; 869 AA.  
 AC  
 XX AA134354;  
 XX  
 DT 25-AUG-1999 (first entry)  
 XX  
 DE Porphyromonas gingivalis protein PG21.  
 XX  
 KM Porphyromonas gingivalis; PG: periodontal disease; gingivitis;  
 KM vaccine; antigenic.  
 XX  
 OS Porphyromonas gingivalis.  
 OS  
 PN WO929870-A1.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 10-DEC-1998; 98WO-AU01023.  
 XX  
 XX  
 PR 04-AUG-1998; 98AU-0005028.  
 PR 10-DEC-1997; 97AU-0000839.  
 PR 31-DEC-1997; 97AU-0001182.  
 PR 30-JAN-1998; 98AU-0001546.  
 PR 10-MAR-1998; 98AU-0002264.  
 PR 09-APR-1998; 98AU-0002911.  
 PR 23-APR-1998; 98AU-0003338.  
 PR 05-MAY-1998; 98AU-0003654.  
 PR 22-MAY-1998; 98AU-0004917.  
 PR 29-JUL-1998; 98AU-0004917.  
 XX  
 XX  
 PA (CSLC-) CSL LTD.  
 XX  
 PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 PI Ross BC, Rothel LJ, Webb EA;  
 XX  
 DR MPI: 1999-385613/32.  
 DR N-PSDB; AA131572.  
 XX  
 PT Antigenic Porphyromonas gingivalis peptides for preventing  
 PT gingivitis  
 XX  
 PS Claim 1: Page 316-318; 588pp; English.  
 XX  
 CC AA13156 to AA131801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AA134318 to  
 CC AA134583. AA131802 to AA131989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation  
 CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 CC  
 SQ Sequence 869 AA;  
 Query Match 89.7%; Score 26; DB 20; Length 869;  
 Best Local Similarity 85.7%; Pred. No. 2.4e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 VMXVAEF 7  
 Db 613 VMPVAEF 619  
 RESULT 5  
 AA056778  
 ID AA056778 standard; Protein; 75 AA.  
 AC  
 XX AA056778;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #17674.  
 XX  
 KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KM dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 OS  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR MPI: 2001-616774/771.  
 DR N-PSDB; AAS59578.  
 XX  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 XX  
 PS Example 1; SEQ ID No 17973; 1069pp; English.  
 XX  
 CC Sequences AA039105-AA068017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting the  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pot\_sequences.  
 CC  
 SQ Sequence 75 AA;  
 Query Match 82.8%; Score 24; DB 22; Length 75;  
 Best Local Similarity 71.4%; Pred. No. 56;

Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 VMXVAEF 7  
 1 1 1 1 1  
 Db 1 VMXVAEF 7

## RESULT 6

AAM41984  
 ID AAM41984 standard; Protein: 132 AA.

AC AAM41984;

DE 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6915.

XX Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokineic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI61140.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX Example 2; SEQ ID NO 6915; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytosstatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokineic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX Sequence 132 AA;

Query Match 82.8%; Score 24; DB 22; Length 132;  
 Best Local Similarity 71.4%; Pred. No. 1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VMXVAEF 7  
 1 1 1 1 1  
 Db 64 LMXVAEF 70

## RESULT 7

AAM41985  
 ID AAM41985 standard; Protein: 132 AA.

AC AAM41985;

DE 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6916.

XX Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokineic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI61141.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX Example 2; SEQ ID NO 6916; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytosstatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokineic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed

CC specification.  
CQ Sequence 132 AA;  
SQ  
Query Match 82.8%; Score 24; DB 22; Length 132;  
Best Local Similarity 71.4%; Pred. No. 1e+02; 1;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VMXVAEF 7  
: | | | |  
DB 64 LMVVAEF 70  
RESULT 8  
AAM41986  
ID AAM41986 standard; Protein; 132 AA.  
AC AAM41986;  
DT 22-OCT-2001 (first entry)  
DE Human polypeptide SEQ ID NO 6917.  
XX  
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukemia.  
XX  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
PD 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX (HSE-) HSEQ INC.  
XX  
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
XX N-PSDB: AA161142.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
XX  
XX Example 2; SEQ ID NO 6917; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with noctropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
CQ Sequence 132 AA;  
SQ  
Query Match 82.8%; Score 24; DB 22; Length 132;  
Best Local Similarity 71.4%; Pred. No. 1e+02; 1;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 VMXVAEF 7  
: | | | |  
DB 64 LMVVAEF 70  
RESULT 9  
AAV85820  
ID AAV85820 standard; Protein; 225 AA.  
AC AAV85820;  
DT 10-APR-2000 (first entry)  
DE S. pneumoniae derived protein #29.  
XX  
XX Treatment; prevention; disease; diagnosis; gene therapy; screening;  
XX bacterial; antimicrobial; antibiotic; pathogenesis; infection.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX WO9806734-A1.  
XX  
XX 19-FEB-1998.  
XX  
XX 15-AUG-1997; 97WO-US14436.  
XX  
XX 16-AUG-1996; 96US-0024022.  
XX  
XX (SMIK) SMITHKLINE BEECHAM CORP.  
XX  
XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
XX Stodola RK;  
XX  
XX WPI: 1998-159452/14.  
XX N-PSDB: AA296199.  
XX  
XX Streptococcus pneumoniae proteins and related DNA - useful for  
XX screening compounds for antibacterial activity  
XX  
XX Claim 5; Page 338-339; 640pp; English.  
XX  
XX This invention describes novel isolated Streptococcus pneumoniae  
XX polynucleotides (see AA296173-296494) and their encoded proteins (see  
XX AA85792-Y86182). The DNA, vectors and host cells described in the  
XX method of the invention are useful for the recombinant expression of  
XX polypeptides. The polypeptides are useful for treatment or prevention of  
XX disease, or diagnosis of disease related to expression or activity of  
XX such a polypeptide. They can also be used to screen for compounds which  
XX interact with and inhibit or activate such a polypeptide. The  
XX polypeptides (or DNA encoding them, via gene therapy) are also useful  
XX for inducing an immunological response in a mammal. The antagonists are  
XX useful to inhibit such bacterial polypeptides. The polypeptides are  
XX particularly useful to identify antimicrobial compounds and antibiotics.  
XX They are also useful to determine their role in pathogenesis of  
XX infection, dysfunction and disease.  
XX  
XX Sequence 225 AA;  
SQ  
Query Match 82.8%; Score 24; DB 19; Length 225;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



DE Novel human diagnostic protein #18325.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS  
 XX MO200175067-A2.  
 PN  
 XX  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS82521.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID NO 48693; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 286 AA:  
 Query Match 82.8%; Score 24; DB 22; Length 286;  
 Best Local Similarity 71.4%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 VMXVAEF 7  
 |||||  
 Db 23 VMOVAQF 29  
 |||||  
 RESULT 13  
 AAU37753  
 ID AAU37753 standard; Protein; 335 AA.  
 AC AAU37753;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Streptococcus pneumoniae cellular proliferation protein #182.  
 XX

KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX Streptococcus pneumoniae.  
 OS  
 XX MO200170955-A2.  
 PN  
 XX  
 PD 27-SEP-2001.  
 PF 21-MAR-2001; 2001WO-US09180.  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyckind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI: 2001-611495/70.  
 DR N-PSDB; AAS55612.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 PS Example 3; Seq ID NO 13346; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 335 AA:  
 Query Match 82.8%; Score 24; DB 22; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 VMXVAEF 7  
 |||||  
 Db 224 VMSNAEF 230  
 |||||  
 RESULT 14  
 AAG20945  
 ID AAG20945 standard; Protein; 361 AA.  
 AC AAG20945;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 23321.  
 XX

KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127452.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130892.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 20-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 24-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 27-MAY-1999; 99US-0136782.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0138094.  
PR 08-JUN-1999; 99US-0138840.  
PR 10-JUN-1999; 99US-0138847.  
PR 10-JUN-1999; 99US-0139119.  
PR 14-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 16-JUN-1999; 99US-0139492.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139465.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161400.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 82.8%; Score 24; DB 21; Length 361;  
Best Local Similarity 71.4%; Pred. No. 3.1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
11:111  
Db 31 VMNMAEF 37

RESULT 15  
ID AAG24458  
ID AAG24458 standard; Protein; 361 AA.

AC AAG24458;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 28136.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.  
PN

XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PF 05-MAR-1999; 99US-0123160.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 29-MAR-1999; 99US-0126264.  
PR 25-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0136624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 82.8%; Score 24; DB 21; Length 361;  
Best Local Similarity 71.4%; Pred. No. 3,1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
| | : | | |  
DB 31 VMNMAEF 37

Search completed: October 30, 2002, 12:27:07  
Job time : 4.80344 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 1.34152 Seconds

(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-78

Perfect score: 29

Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : PIR\_71:\*

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	96.6	2638	1 A42545	genome polyprotein
2	27	93.1	802	2 T05596	probable potassium
3	26	89.7	223	2 E64205	deoxyribose-phosph
4	26	89.7	224	2 S02216	deoxyribose-phosph
5	26	89.7	324	2 B84452	probable steroid s
6	25	86.2	201	2 F69988	hypothetical prote
7	25	86.2	632	2 T38126	probable electron
8	24	82.8	145	1 B41715	ribosomal protein
9	24	82.8	195	2 C97255	thymidine kinase (
10	24	82.8	236	2 B70728	hypothetical prote
11	24	82.8	248	2 T18135	hypothetical prote
12	24	82.8	249	2 G81693	serine/threonine p
13	24	82.8	335	2 F95103	6-phosphofructokin
14	24	82.8	335	2 D97971	6-phosphofructokin
15	24	82.8	367	2 F97010	fusion, chorismate
16	24	82.8	484	2 E86416	unknown protein, 3
17	24	82.8	565	2 B72660	probable type II D
18	24	82.8	582	2 E71052	hypothetical prote
19	24	82.8	584	2 F75090	archaeosine trna-r
20	24	82.8	754	2 T25551	hypothetical prote
21	24	82.8	919	2 T29581	hypothetical prote
22	24	82.8	921	2 H69643	isoleucine--trna 1
23	24	82.8	921	2 A83968	isoleucyl-trna syn
24	24	82.8	1122	2 T42400	Eph receptor tyros
25	23	79.3	89	2 T15018	hypothetical prote
26	23	79.3	96	2 AB3053	hypothetical prote
27	23	79.3	112	2 E90987	hypothetical prote
28	23	79.3	112	2 H85832	hypothetical prote
29	23	79.3	143	2 H69515	riboflavin synthas

30	23	79.3	157	2 G00016	FGF-receptor - com
31	23	79.3	162	2 G86842	shikimate kinase (
32	23	79.3	169	2 AB2739	acetyltransferase
33	23	79.3	172	2 S27019	fibroblast growth
34	23	79.3	172	2 S27020	fibroblast growth
35	23	79.3	172	2 S27021	fibroblast growth
36	23	79.3	172	2 S27022	fibroblast growth
37	23	79.3	187	2 S16314	photosynthetic rea
38	23	79.3	191	2 A82392	transcription regu
39	23	79.3	209	2 H97519	hypothetical prote
40	23	79.3	248	2 D81436	probable tomb tran
41	23	79.3	254	1 BVB953	phosphomannomutase
42	23	79.3	275	2 S28749	NADH dehydrogenase
43	23	79.3	291	2 G69479	methionyl aminopep
44	23	79.3	370	2 F36819	C14 protein - rabb
45	23	79.3	373	2 AC0253	ribonuclease III (

## ALIGNMENTS

```

RESULT 1
A42545 genome polyprotein - Langat virus (strain TP21) (fragment)
N:Contains: nonstructural protein NS1; nonstructural protein NS2a; nonstructural prote
NS5
C:Species: Langat virus
C:Date: 30-Sep-1993 #sequence.revision 30-Sep-1993 #text.change 19-Jan-2001
C:Accession: A42545; A61409; G61409
R:Iacono-Connors, L.C.; Schmaljohn, C.S.
Virology 188, 875-880, 1992
A:Title: Cloning and sequence analysis of the genes encoding the nonstructural protei
A:Reference number: A42545; MUID:92263794
A:Accession: A42545
A:Molecule type: genomic RNA
A:Residues: 1-2638 <IAC>
A:Cross-references: GB:S35365; NID:g249315; PIDN:AAB22165.1; PID:g249316
R:Gutrahoo, F.; Helms, F.X.; Mandl, C.W.; Holzmann, H.; Kunz, C.; Gresikova, M.
J. Gen. Virol. 72, 333-338, 1991
A:Title: The relationship between the flaviviruses Skatlica and Langat as revealed by
A:Reference number: A61409; MUID:91132129
A:Accession: A61409
A>Status: not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 319-337 <GUI>
A:Accession: G61409
A>Status: not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 877-994 <GU2>
C:Superfamily: yellow fever virus genome polyprotein
C:Keywords: glycoprotein; nonstructural protein; nucleotide binding; P-loop; polyprot
F:1352/Product: nonstructural protein NS1 #status predicted <NS1>
F:353-562/Product: nonstructural protein NS2a #status predicted <NS2>
F:583-713/Product: nonstructural protein NS2b #status predicted <NS2>
F:714-1334/Product: nonstructural protein NS3 #status predicted <NS3>
F:912-919/Region: nucleotide-binding motif A (P-loop)
F:1335-1483/Product: nonstructural protein NS4 #status predicted <NS4>
F:1484-1733/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1736-2638/Product: nonstructural protein NS5 #status predicted <NS5>
F:85,207,223,873,1212,1671,1950/Binding site: carbohydrate (asn) (covalent) #status p
Query Match 96.6%; Score 28; DB 1; Length 2638;
Best Local Similarity 85.7%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 VMXVAEF 7
Db 153 VMTVAEF 159
RESULT 2
T05596 probable potassium transport protein F9D16.110 - Arabidopsis thaliana

```

C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 18-Aug-2000  
 C:Accession: T05596  
 R:Devan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohsels, J.; Mewes, H.W.; Mayer, K.F.  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: 215419  
 A:Accession: T05596  
 A:Molecule type: DNA  
 A:Residues: 1-802 <BPV>  
 A:Cross-references: EMBL:AL035394  
 A:Experimental source: cultivar Columbia; BAC clone F9D16  
 C:Genetics:  
 A:Map position: 4  
 A:Insertions: 1/3; 16/3; 102/1; 185/1; 209/1; 296/1; 313/3; 352/1; 437/1  
 A>Note: F9D16.110  
 C:Superfamily: barley probable potassium transport protein HMK1  
 C:Keywords: ion transport

Query Match 93.1%; Score 27; DB 2; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 43;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 ||:||||  
 Db 629 VM5IAEF 635

RESULT 3  
 E64205  
 deoxyribose-phosphate aldolase (EC 4.1.2.4) - Mycoplasma genitalium  
 C:Species: Mycoplasma genitalium  
 C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 21-Jul-2000  
 A:Accession: E64205  
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;  
 M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.  
 C.A.; Venter, J.C.  
 Science 270, 397-403, 1995  
 A:Title: The minimal gene complement of Mycoplasma genitalium.  
 A:Reference number: A64200; MUID:96026346  
 A:Accession: E64205  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-223 <TIGR>  
 A:Cross-references: GB:U93684; GB:L43967; NID:93844650; PIDN:AAC71266.1; PID:91045723;  
 A:Experimental source: strain G-37  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: deoxyribose-phosphate aldolase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 89.7%; Score 26; DB 2; Length 223;  
 Best Local Similarity 71.4%; Pred. No. 21;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 ||:||||  
 Db 93 VM5IAEF 99

RESULT 4  
 S02216  
 deoxyribose-phosphate aldolase (EC 4.1.2.4) - Mycoplasma pneumoniae  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 07-Dec-1999  
 A:Accession: S02216; S73417  
 R:Loechel, S.; Inamine, J.M.; Hu, P.C.  
 Nucleic Acids Res. 17, 801, 1989  
 A:Title: Nucleotide sequence of the deoc gene of Mycoplasma pneumoniae.  
 A:Reference number: S02216; MUID:89128453  
 A:Accession: S02216  
 A:Molecule type: DNA  
 A:Residues: 1-224 <LOE>

A:Cross-references: EMBL:X13544; NID:944480; PIDN:CA31897.1; PID:944481  
 R:Himmelfreid, R.; Hildbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S73327; MUID:97105885  
 A:Accession: S73417  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-224 <HIM>  
 A:Cross-references: EMBL:AE000011; GB:U00089; NID:91673740; PIDN:AAB95739.1; PID:9167  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C:Genetics:  
 A:Gene: deoc  
 A:Genetic code: SGC3  
 C:Superfamily: deoxyribose-phosphate aldolase  
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 89.7%; Score 26; DB 2; Length 224;  
 Best Local Similarity 71.4%; Pred. No. 22;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 ||:||||  
 Db 93 VM5IAEF 99

RESULT 5  
 B84452  
 probable steroid sulfotransferase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 A:Accession: B84452  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanNieuwen, S.E.; Umayam, L.; Tallon,  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: B84420; MUID:20083487  
 A:Accession: B84452  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-324 <STO>  
 A:Cross-references: GB:AE002093; NID:94406768; PIDN:AAD20079.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g03770  
 A:Map position: 2  
 C:Superfamily: alcohol sulfotransferase

Query Match 89.7%; Score 26; DB 2; Length 324;  
 Best Local Similarity 71.4%; Pred. No. 31;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 ||:||||  
 Db 233 VM5IAEF 239

RESULT 6  
 F69988  
 hypothetical protein ytbQ - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 A:Accession: F69988  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C.; Bron, S.; Brouillette, S.; Bursch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;  
 A.; Ehrlich, S.D.; Emerson, P.R.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal  
 lech, J.; Harwood, C.R.; Henaut, A.; Hildbert, H.; Holstappel, S.; Hosono, S.; Hullo, M  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardino  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scioffone, F.; Sekiuchi, J.; Sekowska, A.; Serot  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: F69988  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-201 <KUN>  
A:Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14996.1; PID:e1185891;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ytbq

Query Match  
Best Local Similarity 86.2%; Score 25; DB 2; Length 201;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
1 : |||  
DB 23 IMDIAEF 29

RESULT 7  
T38126  
probable electron transfer flavoprotein precursor - fission yeast (*Schizosaccharomyces p*  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T38126  
R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1997  
A:Reference number: Z21772  
A:Accession: T38126  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-632 <BAD>  
A:Cross-references: EMBL:Z95334; PIDN:CAB08598.1; GSPDB:GN00066; SPDB:SPAC2068.04c  
A:Experimental source: strain 972h-; cosmid c2068  
C:Genetics:  
A:Gene: SPDB:SPAC2068.04c  
A:Map position: 1

Query Match  
Best Local Similarity 86.2%; Score 25; DB 2; Length 632;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
1 : |||  
DB 193 VMSIAEF 199

RESULT 8  
B41715  
ribosomal protein L13 [similarity] - *Halococcus marismortui*  
C:Species: *Halococcus marismortui*  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: B41715  
R:Kroemer, W.D.; Arndt, E.  
J. Biol. Chem. 266, 24573-24579, 1991  
A:Title: Halobacterial S9 operon. Three ribosomal protein genes are cotranscribed with a  
bacterial marismortui.  
A:Reference number: A41715; MUID:92105119  
A:Accession: B41715  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <KRO>  
A:Cross-references: GB:M76567; NID:g148775; PIDN:AA73097.1; PID:g148777  
C:Superfamily: *Escherichia coli* ribosomal protein L13

Query Match  
Best Local Similarity 82.8%; Score 24; DB 1; Length 145;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MXVAEF 7  
1 : |||  
DB 1 MCVAEF 6

RESULT 9  
C97255  
thymidine kinase (EC 2.7.1.21) [similarity] - *Clostridium acetobutylicum*  
C:Species: *Clostridium acetobutylicum*  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 31-Dec-2001  
C:Accession: C97255  
R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97255  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK0830.1; PID:g15025935; GSPDB:GN00168  
A:Experimental source: *Clostridium acetobutylicum* ATCC824  
C:Genetics:  
A:Gene: CAC2887  
C:Superfamily: thymidine kinase  
C:Keywords: phosphotransferase

Query Match  
Best Local Similarity 82.8%; Score 24; DB 2; Length 195;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
1 : |||  
DB 131 LMAIAEF 137

RESULT 10  
B70728  
hypothetical protein RV2558 - *Mycobacterium tuberculosis* (strain H37RV)  
C:Species: *Mycobacterium tuberculosis*  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: B70728  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Rajandream, M.A.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd,  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70728  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-236 <COL>  
A:Cross-references: GB:Z77250; GB:AL123456; NID:g3261617; PIDN:CAB01046.1; PID:e25533  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV2558

Query Match  
Best Local Similarity 82.8%; Score 24; DB 2; Length 236;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
1 : |||  
DB 217 VLDVAEF 223

RESULT 11  
T18315  
hypothetical protein L7610.5 - *Leishmania major*  
C:Species: *Leishmania major*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18315  
 R:Oliver, K.; Murphy, L.; Quail, M.; Lawson, D.; Harris, D.; Rajandream, M.; Ivens, A.;  
 submitted to the EMBL Data Library, May 1999  
 A:Reference number: 218876  
 A:Accession: T18315  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-248 <OLJ>  
 A:Cross-references: EMBL:AL034356; NID:e1371876; PID:e1371560; PIDN:CAA2241.1  
 C:Genetics:  
 A:Note: l7610.5

Query Match 82.8%; Score 24; DB 2; Length 248;  
 Best Local Similarity 71.4%; Pred. No. 83;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 || ||||  
 Db 196 LMRVAEF 202

## RESULT 12

G81693  
 serine/threonine protein phosphatase, probable TC0530 [imported] - Chlamydia muridarum  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
 C:Accession: G81693  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255

A:Accession: G81693  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-249 <TEP>  
 A:Cross-references: GB:AE002321; GB:AE002160; NID:g7190560; PIDN:AAF9372.1; PID:g719057  
 A:Experimental source: strain Nigg (Mopn)  
 C:Genetics:  
 A:Gene: TC0530  
 C:Superfamily: conserved hypothetical protein y100; conserved hypothetical protein y100

Query Match 82.8%; Score 24; DB 2; Length 249;  
 Best Local Similarity 71.4%; Pred. No. 84;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 || ||||  
 Db 1 MMIVAEF 7

## RESULT 13

F95103  
 6-phosphofructokinase [imported] - Streptococcus pneumoniae (strain TIGR4)  
 C:Species: Streptococcus pneumoniae  
 C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
 C:Accession: F95103  
 R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held,  
 om, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,  
 nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
 A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95103  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-335 <KUR>  
 A:Cross-references: GB:AE005672; PIDN:AAK75023.1; PID:g1497371; GSPDB:GN00164; TIGR:SP4  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SP0896  
 C:Superfamily: 6-phosphofructokinase; 6-phosphofructokinase 1 homology

Query Match 82.8%; Score 24; DB 2; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 11e+02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 || ||||  
 Db 224 VMSAAEF 230

## RESULT 14

D97971  
 6-phosphofructokinase (EC 2.7.1.11) [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C:Accession: D97971  
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: D97971  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-335 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAK99600.1; PID:g15458395; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: pfkA  
 C:Superfamily: 6-phosphofructokinase; 6-phosphofructokinase 1 homology

C:Keywords: phosphotransferase  
 Query Match 82.8%; Score 24; DB 2; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 || ||||  
 Db 224 VMSAAEF 230

## RESULT 15

F97010  
 fusion, chorismate mutase and splikimate 5-dehydrogenase [imported] - Clostridium acet  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: F97010  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L  
 .; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: F97010  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-367 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK78873.1; PID:g15023795; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC0897

Query Match 82.8%; Score 24; DB 2; Length 367;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
 || ||||  
 Db 29 VMEVAEF 35

Search completed: October 30, 2002, 12:31:31  
 Job time : 4.34152 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:10 ; Search time 0.601966 Seconds  
(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-78  
Perfect score: 29  
Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : SWISSPROT\_40.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	96.6	3414	1	P29837 1 genome po
2	26	89.7	223	1	P47296 MYCOPASMA
3	26	89.7	224	1	P09924 MYCOPASMA
4	25	86.2	253	1	P53560 bacillus su
5	25	86.2	632	1	P87111 s probable
6	24	82.8	145	1	P29198 halococcula
7	24	82.8	157	1	P49065 taraxacum o
8	24	82.8	236	1	P50740 MYCOPASMA
9	24	82.8	921	1	P04547 bacillus su
10	24	82.8	921	1	P04547 bacillus su
11	24	82.8	971	1	P04547 bacillus su
12	24	82.8	1095	1	P04547 bacillus su
13	23	79.3	143	1	P8195 mus musculu
14	23	79.3	162	1	P08195 mus musculu
15	23	79.3	187	1	P08195 mus musculu
16	23	79.3	254	1	P08195 mus musculu
17	23	79.3	275	1	P08195 mus musculu
18	23	79.3	291	1	P08195 mus musculu
19	23	79.3	299	1	P08195 mus musculu
20	23	79.3	332	1	P08195 mus musculu
21	23	79.3	632	1	P08195 mus musculu
22	23	79.3	654	1	P08195 mus musculu
23	23	79.3	773	1	P08195 mus musculu
24	23	79.3	773	1	P08195 mus musculu
25	23	79.3	801	1	P08195 mus musculu
26	23	79.3	801	1	P08195 mus musculu
27	23	79.3	806	1	P08195 mus musculu
28	23	79.3	806	1	P08195 mus musculu
29	23	79.3	808	1	P08195 mus musculu
30	23	79.3	813	1	P08195 mus musculu
31	23	79.3	819	1	P08195 mus musculu
32	23	79.3	821	1	P08195 mus musculu
33	23	79.3	821	1	P08195 mus musculu

34	23	79.3	822	1	FGRL_HUMAN	P11362 homo sapien
35	23	79.3	822	1	FGRL_MOUSE	P16092 mus musculu
36	23	79.3	822	1	FGRL_RAT	P04589 ratu mus
37	23	79.3	823	1	CEK3_CHICK	P18461 gallu galli
38	23	79.3	852	1	VTER_HSV1	P00140 ictaluri h
39	23	79.3	980	1	POLG_LIV	P22338 louping ill
40	23	79.3	1002	1	EPB5_CHICK	P07497 gallu galli
41	23	79.3	1052	1	FGRL_MOUSE	P09147 drosophila
42	23	79.3	1914	1	STCK_EMENT	P00706 emeritella
43	23	79.3	3412	1	POLG_MOUSE	P07720 t genome po
44	23	79.3	3414	1	POLG_MOUSE	P01299 t genome po
45	23	79.3	3414	1	POLG_MOUSE	P14336 t genome po

## ALIGNMENTS

RESULT 1  
ID POLG\_LANVT STANDARD: PRT: 3414 AA.  
AC P29837.  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polypeptide [Contains: Capsid protein C (core protein);  
DE Envelope protein PM; Matrix protein (Envelope protein M); Major  
DE envelope protein E; Nonstructural protein NS1; Nonstructural protein  
DE NS2a; Nonstructural protein NS2B; Helicase/protease (EC 3.4.21.98)  
DE (NS3); Nonstructural protein NS4A; Nonstructural protein NS4B; RNA-  
DE directed RNA polymerase (EC 2.7.7.48) (NS5)].  
DE Langat virus (strain TP21).  
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Flaviviridae;  
OC Flavivirus.  
OX NCBI\_TaxID=31638;  
RN [1]  
RP SEQUENCE OF 1-776 FROM N.A.  
RX MEDLINE=92074260; PubMed=1720591;  
RA Mandl C.W., Iacono-Connors L., Wallner G., Holzmann H., Kunz C.,  
RA Heinz F.X.;  
RT "Sequence of the genes encoding the structural proteins of the low-  
RT virulence tick-borne flaviviruses Langat TP21 and Yelantsev.";  
RT Virology 185:891-895(1991).  
RL [2]  
RL SEQUENCE OF 777-3414 FROM N.A.  
RX MEDLINE=92263794; PubMed=1316684;  
RA Iacono-Connors L.C., Schmaljohn C.S.;  
RT "Cloning and sequence analysis of the genes encoding the  
RT nonstructural proteins of Langat virus and comparative analysis with  
RT other flaviviruses.";  
RL Virology 188:875-880(1992).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF FOUR PEPTIDE BONDS IN THE VIRAL  
CC PRECURSOR POLYPEPTIDE, COMMONLY WITH ASP OR GLU IN THE P6  
CC POSITION, CYS OR THR IN P1 AND SER OR ALA IN P1'.  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
DR EMBL: M73835; AAA02740.1; ALT-TERM.  
DR EMBL: S35365; AAB22165.1; -.  
DR PIR: A41704; A41704.  
DR PIR: A42545; A42545.  
DR HSSP: P14336; 1SVB.  
DR MEROPS: S07.001; -.  
DR InterPro: IPR001410; DEAD.

```

DR InterPro: IPR000069; Flavl_M.
DR InterPro: IPR001157; Flavl_NSI.
DR InterPro: IPR000752; Flavl_NS2A.
DR InterPro: IPR000487; Flavl_NS2B.
DR InterPro: IPR000404; Flavl_NS4A.
DR InterPro: IPR001528; Flavl_NS4B.
DR InterPro: IPR000208; Flavl_NS5.
DR InterPro: IPR001122; Flavl_capsid.
DR InterPro: IPR000336; Flavl_glycoprote.
DR InterPro: IPR001850; Flavl_helicase.
DR InterPro: IPR002535; Flavl_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF01003; Flavl_capsid; 1.
DR Pfam: PF00869; Flavl_glycoprote; 1.
DR Pfam: PF02832; Flavl_glycop_C; 1.
DR Pfam: PF00949; Flavl_helicase; 1.
DR Pfam: PF01004; Flavl_M; 1.
DR Pfam: PF00948; Flavl_NSI; 1.
DR Pfam: PF01005; Flavl_NS2A; 1.
DR Pfam: PF01002; Flavl_NS2B; 1.
DR Pfam: PF01350; Flavl_NS4A; 1.
DR Pfam: PF01348; Flavl_NS4B; 1.
DR Pfam: PF00972; Flavl_NS5; 1.
DR Pfam: PF01570; Flavl_propep; 1.
DR Pfam: PF01728; FtsJ; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF001496; Flavl_NSI; 1.
DR Pfam: PF001556; Flavl_glycoprote; 1.
DR SMART: SMO0490; HelicC; 1.
DR Polyprotein: Glycoprotein; Transferrase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Hydrolyase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein
FT INIT_MBT 1
FT CHAIN 1 112
FT CHAIN 113 205
FT CHAIN 206 280
FT CHAIN 281 776
FT CHAIN 777 1128
FT CHAIN 1129 1358
FT CHAIN 1359 1489
FT CHAIN 1490 2110
FT CHAIN 2111 2259
FT CHAIN 2260 2511
FT CHAIN 2512 3414
FT NP_BIND 1688 1695
FT SITE 1779 1782
FT TRANSMEM 103 119
FT TRANSMEM 262 278
FT TRANSMEM 728 744
FT TRANSMEM 758 774
FT TRANSMEM 283 310
FT DISULFID 340 396
FT DISULFID 354 385
FT DISULFID 372 401
FT DISULFID 466 570
FT DISULFID 587 618
FT CARBOHYD 144 144
FT CARBOHYD 434 434
SQ SEQUENCE 3414 AA; 378017 MW; 59CBFE95DD70D82E CRC64;

```

```

Query Match 96.6%; Score 28; DB 1; Length 3414;
Best Local Similarity 85.7%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 VMXVAEF 7
DB 929 VMYVAEF 935

```

RESULT 2

```

DEOC_MYCGE ID DEOC_MYCGE STANDARD; PRT; 223 AA.
AC P47296;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyribosaldolase)
DE (Deoxyribosaldolase).
GN DEOC OR M6050.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fitchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -I- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC glyceraldehyde 3-phosphate + acetaldehyde.
CC -I- PATHWAY: NUCLEOTIDE AND DEOXYRIBONUCLEOTIDE CATABOLISM.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.
CC DEOC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39684; AAC71266.1; -.
CC TIGR: MG050; -.
DR InterPro: IPR002915; DEOC.
DR InterPro: IPR003009; FNN_enzyme.
DR Pfam: PF01791; DEOC; 1.
KM Lyase; Schiff base; Complete proteome.
FT BINDING 152 152
FT BINDING 152 152
SQ SEQUENCE 223 AA; 24675 MW; 332430231CE99DB0 CRC64;

```

```

Query Match 89.7%; Score 26; DB 1; Length 223;
Best Local Similarity 71.4%; Pred. No. 8;7;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 VMXVAEF 7
DB 93 VMYVAEF 99

```

```

RESULT 3
DEOC_MYCPN ID DEOC_MYCPN STANDARD; PRT; 224 AA.
AC P09924;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyribosaldolase)
DE (Deoxyribosaldolase).
GN DEOC OR MPN063 OR MP091.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]

```



RC SEQUENCE FROM N.A.  
RC STRAIN-ATCC 29342 / M129;  
RX MEDLINE=89128453; PubMed=2492658;  
RA Loechel S., Inamine J.M., Hu P.-C.;  
RT "Nucleotide sequence of the deoc gene of Mycoplasma pneumoniae";  
RL Nucleic Acids Res. 17:801-801(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelfreuch R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RT Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -1- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate - D-  
CC glycerinaldehyde 3-phosphate + acetaldehyde.  
CC -1- PATHWAY: NUCLEOTIDE AND DEOXYRIBONUCLEOTIDE CATABOLISM.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.  
CC DEOC SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X13544; CAA31897.1; -  
DR EMBL: AE000011; AAB95739.1; -  
DR PIR: S02216; S02216.  
DR InterPro: IPR002915; Deco.  
DR Pfam: PF01791; Deco; 1.  
KW Lyase; Schiff base; Complete proteome.  
FT BINDING 152 152 SCHIFF-BASE (BY SIMILARITY).  
SQ SEQUENCE 224 AA; 24878 MW; 73CCE4932E7881F7 CRC64;  
  
Query Match 89.7%; Score 26; DB 1; Length 224;  
Best Local Similarity 71.4%; Pred. No. 8; Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 VNMVAEF 7  
Db 93 VNMIAEF 99  
  
RESULT 4  
YTBO\_BACSU  
ID YTBO\_BACSU STANDARD; PRT: 253 AA.  
AC P53560;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein ytbQ.  
GN YTBO.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96312354; PubMed=8763940;  
RA Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P.,  
RT Petro J.;  
RT "Cloning, sequencing, and characterization of the Bacillus subtilis  
RT biotin biosynthetic operon";  
RL J. Bacteriol. 178:4122-4130(1996).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U51868; AAB17463.1; -  
DR Subtilisin; BG11787; ytbQ.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 253 AA; 28233 MW; 572CC4681F04C650 CRC64;  
  
Query Match 86.2%; Score 25; DB 1; Length 253;  
Best Local Similarity 57.1%; Pred. No. 18; Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 VNMVAEF 7  
Db 75 IMDIAEF 81  
  
RESULT 5  
ETFD\_SCHPO  
ID ETFD\_SCHPO STANDARD; PRT: 632 AA.  
AC P87111;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Probable electron transfer flavoprotein-ubiquinone oxidoreductase,  
DE mitochondrial precursor (Ec 1.5.5.1) (ETF-OO) (ETF-ubiquinone  
DE oxidoreductase) (ETF dehydrogenase) (Electron-transferring-  
DE flavoprotein dehydrogenase).  
GN SPAC2068.04C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RL Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACCEPTS ELECTRONS FROM ETF AND REDUCES UBIQUINONE (BY  
CC SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: Reduced ETF + ubiquinone -> ETF + ubiquinol.  
CC -1- COFACTOR: FAD AND A 4FE-4S CLUSTER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By  
CC similarity).  
CC -1- SIMILARITY: BELONGS TO THE ETF-OO / FIXC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z95334; CAB08598.1; -  
DR Oxidoreductase; Electron transport; Flavoprotein; FAD; Iron-sulfur;  
KW 4Fe-4S; Mitochondrion; Transist peptide; Ubiquinone.  
FT TRANSIT 1 ?  
FT CHAIN ? 632  
FT NE\_BIND 93 107  
FT METAL 575 575  
FT METAL 601 601  
FT METAL 604 604  
FT METAL 607 607  
SQ SEQUENCE 632 AA; 69472 MW; 1B0F22374E33771B CRC64;  
  
Query Match 86.2%; Score 25; DB 1; Length 632;  
Best Local Similarity 71.4%; Pred. No. 45; Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MVXAEF 7  
 DB 193 VMSIAEF 199

RESULT 6  
 RL13\_HALMA STANDARD; PRT: 145 AA.  
 AC P29198;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 50S ribosomal protein L13p (Hmal13).  
 GN RPL13p.  
 OS Halorcula marismortui (Halobacterium marismortui).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.  
 NCBI\_TaxID=2238;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92105119; PubMed=1840597;  
 RA Kroeber W.J., Arndt E.;  
 RT "Halobacterial 5S operon. Three ribosomal protein genes are  
 RT cotranscribed with genes encoding a tRNA (Leu), the enolase, and a  
 RT putative membrane protein in the archaeobacterium Haloarcula  
 RT (Halobacterium) marismortui.";  
 RT J. Biol. Chem. 266:24573-24579(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RX STRAIN=ATCC 43049; PubMed=10937989;  
 RA Ban N., Nissen P., Hansen J., Moore P.B., Steitz T.A.;  
 RT "The complete atomic structure of the large ribosomal subunit at 2.4  
 RT A resolution.";  
 RT Science 289:905-920(2000).  
 CC -1- SIMILARITY: BELONGS TO THE L13P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M76567; AAA73097.1; -  
 CC DR PIR: B41715; B41715.  
 CC DR PDB: 1FEK; 14-AUG-00.  
 CC DR InterPro: IPR001074; Ribosomal\_L13.  
 CC DR Pfam: PF00572; Ribosomal\_L13.  
 CC DR ProDom: PD001791; Ribosomal\_L13; 1.  
 CC DR PROSITE: PS00783; RIBOSOMAL\_L13; 1.  
 CC KW Ribosomal protein; 3D-structure.  
 CC RN SEQUENCE 145 AA; 16228 MW; 069CE66662AE3BC CRC64;  
 SO

Query Match 82.8%; Score 24; DB 1; Length 145;  
 Best Local Similarity 83.3%; Pred. No. 20;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MVXAEF 7  
 DB 1 MSVAEF 6

RESULT 7  
 RAP\_TAROF STANDARD; PRT: 157 AA.  
 ID RAP\_TAROF  
 AC 049065;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Root allergen protein (RAP).  
 OS Taraxacum officinale (Common dandelion).  
 SO

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 CC Taraxacum.  
 CC NCBI\_TaxID=50225;  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE=Root;  
 CC RA Xu X.-Y., Bewley J.D., Greenwood J.S.;  
 CC Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE BETV1 FAMILY OF PATHOGENESIS-RELATED  
 CC PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF036931; AAB92255.1; -  
 CC DR HSSP: P15494; 1BVL.  
 CC DR HSSP: P15494; 1BVL.  
 CC DR HSSP: P15494; 1BVL.  
 CC DR InterPro: IPR000916; Bet\_v\_1.  
 CC DR Pfam: PF00407; Bet\_v\_1; 1.  
 CC DR PRINTS: PR00634; BETALLERGEN.  
 CC DR PRODOM: PD000531; Bet\_v\_1; 1.  
 CC DR PROSITE: PS00451; PATHOGENESIS\_BETV1; 1.  
 CC KW Allergen.  
 CC RN SEQUENCE 157 AA; 17040 MW; 5892AB8593A8A7E0 CRC64;  
 SO

Query Match 82.8%; Score 24; DB 1; Length 157;  
 Best Local Similarity 83.3%; Pred. No. 21;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MVXAEF 7  
 DB 1 MVXAEF 6

RESULT 8  
 YP58\_MYCTU STANDARD; PRT: 236 AA.  
 ID YP58\_MYCTU  
 AC 050740;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 25.7 kDa protein RV2558.  
 GN RV2558 OR MT2635 OR MTCY9C4.10C.  
 OS Mycobacterium tuberculosis.  
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 CC NCBI\_TaxID=1773;  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=H37RV;  
 CC MEDLINE=98295987; PubMed=9634230;  
 CC Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 CC Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 CC Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 CC Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 CC Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 CC Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 CC Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 CC Sulston J.E., Taylor K., Whitehead S., Whitehead B.G.;  
 CC "Deciphering the biology of Mycobacterium tuberculosis from the  
 CC complete genome sequence.";  
 CC Nature 393:537-544(1998).  
 CC [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=CDC 1551 / Oshkosh;  
 CC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains";  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV2557.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z77250; CAB01046.1; -  
DR EMBL: AE007098; AAK46947.1; -  
DR TIGR: MT635; -  
DR TubercuList: RV2558; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 236 AA; 25718 MW; 13E3B049DBF79C6B CRC64;

Query Match  
Best Local Similarity 82.8%; Score 24; DB 1; Length 236;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAF 7  
ID 1: ||||  
DB 217 VLDVAF 223

RESULT 9  
TP6B\_AERPE  
ID TP6B\_AERPE STANDARD: PRT; 565 AA.  
AC O9Y64;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Type II DNA topoisomerase VI subunit B (EC 5.99.1.3).  
GN TOPEB OR APE0706.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;  
OC Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankal A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
crenarchaeon, Aeropyrum pernix K1";  
RL DNA Res. 6:83-101(1999).  
CC -1- FUNCTION: RELAXES BOTH POSITIVE AND NEGATIVE SUPERURNS AND  
EXHIBITS A STRONG DEACETYLASE ACTIVITY. THE B SUBUNIT BINDS ATP (BY  
SIMILARITY)  
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
of double-stranded DNA.  
CC -1- SUBUNIT: HETEROTETRAMER OF TWO SUBUNITS A AND TWO SUBUNITS B (BY  
SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AP000060; BAA/9682.1; -  
DR InterPro: IPR003594; HATPase\_C.  
DR InterPro: IPR004359; HIS\_KIN\_sig.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR SMART: SM00387; HATPase\_C; 1.  
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Complete proteome.  
SQ SEQUENCE 565 AA; 64096 MW; E4A3DFC5E53D508 CRC64;

Query Match  
Best Local Similarity 82.8%; Score 24; DB 1; Length 565;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MXYAER 7  
ID 1: ||||  
DB 28 MSVAF 33

RESULT 10  
SYL\_BACSU  
ID SYL\_BACSU STANDARD: PRT; 921 AA.  
AC Q54547; P71022; O31730;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase)  
DE (IleRS).  
GN ILRS.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RP SEQUENCE OF 1-121 FROM N.A.  
RC STRAIN-168;  
RA Stewart G.C., Cha J.H.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RP SEQUENCE OF 733-921 FROM N.A.  
RC STRAIN-168;  
RA Pragai Z., Tjalsma H., Bolhuis A., van Dijk J.M., Venema G.,  
RA Bron S.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: ATP + L-Isoleucine + tRNA(Ile) = AMP +  
diphosphate + L-Isoleucyl-tRNA(Ile).  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z69112; CAB1417.1; -  
DR EMBL: U60901; AAB45280.1; -  
DR EMBL: U48870; AAB57764.1; -  
DR HSSP: P41972; 1003.  
DR Subtilist: BG11792; IleS.  
DR InterPro: IPR002300; tRNA-synt\_1a.  
DR InterPro: IPR001412; tRNA-synt\_1.  
DR InterPro: IPR002301; tRNA-synt\_1le.  
DR Pfam: PF00133; tRNA-synt\_1; 1.

```

DR PRINTS: PR00984; TRNASTYHILE.
DR PROSITE; PS00178; AA-TRNA-LIGASE-I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Complete proteome.
FT SITE 57 67 "HIGH" REGION.
FT BINDING 594 598 "KMSK" REGION.
FT CONFUCT 744 744 S -> F (IN REF. 3).
SQ SEQUENCE 921 AA; 104784 MW; 011F31E5F7460D43 CRC64;

Query Match 82.8%; Score 24; DB 1; Length 921;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MXVAEF 7
DB 116 MSVAEF 121

RESULT 11
AMPN_HAECC STANDARD; PRT; 971 AA.
ID AMPN_HAECC
AC 010737;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Aminoacyl-tRNA synthetase (Membrane
glycoprotein H11).
OS Haemonchus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OC NCBI_Taxid=6289;
RX MEDLINE-97273974; PubMed-9128148;
RA Smith T.S., Graham M., Mann E.A., Newton S.E., Knox D.P.,
RA Coadwell W.J., McMichael-Phillips D., Smith W.D.,
RA Oliver J.J.;
RT "Cloning and characterization of a microsomal aminoacyl-tRNA synthetase from the
RT intestine of the nematode Haemonchus contortus."
RL Biochim. Biophys. Acta 1338:295-306(1997).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid,
CC preferably a neutral or hydrophobic one, from a polypeptide.
CC Aminoacyl-arylamides are poor substrates.
CC -1- COFACTOR: BINDS 1 ZINC ION.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X94187; CAA63897.1; -
DR MEROPS; M01.001; -
DR InterPro; IPR001930; Aladiptase.
DR InterPro; IPR00130; Zn_MTPeptide.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolyase; Metalloprotease; Aminoacyl-tRNA synthetase; Zinc; Glycoprotein;
KW Transmembrane; Signal-anchor.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 38 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 39 971 EXTRACELLULAR (POTENTIAL).
FT METAL 378 378 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 379 379 BY SIMILARITY.
FT METAL 382 382 ZINC (CATALYTIC) (BY SIMILARITY).

```

```

FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 465 465 PROTON DONOR (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 971 AA; 110542 MW; 95C6A92B5CCA227C CRC64;

Query Match 82.8%; Score 24; DB 1; Length 971;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7
DB 541 VISVAEF 547

RESULT 12
AT9B_MOUSE STANDARD; PRT; 1095 AA.
ID AT9B_MOUSE
AC P68195;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Potential phospholipid-transporting ATPase IIb (EC 3.6.3.1).
GN ATP9B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_Taxid=10090;
RX MEDLINE-20473714; PubMed-11015572;
RA Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,
RA Hacker C., Pyke S., Newman J.T., Nakamshi Y., Ando H., Weinstock D.,
RA Williamson P., Schlegel R.A.;
RT "Differential expression of putative transblayer amphipath
RT transporters."
RL Physiol. Genomics 1:139-150(1999).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES EXCEPT SPLEEN AND
CC MUSCLE. MOST ABUNDANT IN TESTIS. ALSO DETECTED IN FETAL TISSUES.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (EI-E2
CC ATPASES). SUBFAMILY IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF155913; AAF08476.1; -
DR MGD; MGI:1354757; ATP9B.
DR InterPro; IPR001757; EI-E2-ATPase.
DR InterPro; IPR001454; Hydrolyase.
DR Pfam; PF00122; EI-E2-ATPase; 1.
DR Pfam; PF00702; Hydrolyase; 1.
DR PRINTS; PR00119; CATAATPASE.
DR PROSITE; PS00154; ATPASE_EI_E2; 1.
KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW Multigene family.
FT DOMAIN 1 94 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 95 116 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 117 121 POTENTIAL.
FT TRANSMEM 122 144 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 145 328 POTENTIAL.
FT TRANSMEM 329 349 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 350 357 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 358 379 POTENTIAL.

```

```
FT DOMAIN 380 878 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 879 899 POTENTIAL.
FT DOMAIN 900 911 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 912 930 POTENTIAL.
FT DOMAIN 931 960 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 961 979 POTENTIAL.
FT DOMAIN 980 986 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 987 1009 POTENTIAL.
FT DOMAIN 1010 1015 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 1016 1036 POTENTIAL.
FT DOMAIN 1037 1053 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 1054 1078 POTENTIAL.
FT DOMAIN 1079 1095 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 416 416 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 822 822 MAGNESIUM (BY SIMILARITY).
FT METAL 826 826 MAGNESIUM (BY SIMILARITY).
SO SEQUENCE 1095 AA; 123506 MW; 76D37684241D17E6 CRC64;

Query Match
Best Local Similarity 82.8%; Score 24; DB 1; Length 1095;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MVXAEF 7
DB 1016 LMVVAEF 1022

RESULT 13
RISB_ARCFU STANDARD: PRT: 143 AA.
AC 028152;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 41, Last annotation update)
DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
DE (Lumazine synthase) (Riboflavin synthase beta chain).
GN RIBH OR RIBE OR AF2128.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C.,
RA Kikness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Welch C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Goeyne J.D., Weidman J.F., McDonald L., Uitterback T.,
RA Cotton M.D., Spriggs T., Arltach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinone and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinone with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1-D-ribityl)lumazine =
CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -1- PATHWAY: Riboflavin synthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE DMRL SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
CC EMBL: AE000957; AAB89124.1; -.
CC DR HSSP: P11998; IRVV.
CC DR TIGR: AF2128; -.
CC DR InterPro: IPR002180; DMRL_synthase.
CC DR Pfam: PF00885; DMRL_synthase; 1.
CC KW Riboflavin biosynthesis; Transferase; Complete proteome.
CC SO SEQUENCE 143 AA; 15645 MW; BBA3141206905410 CRC64;

Query Match
Best Local Similarity 79.3%; Score 23; DB 1; Length 143;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MVXAEF 7
DB 8 MVXAEF 13

RESULT 14
AROK_LACLA STANDARD: PRT: 162 AA.
ID AROK_LACLA
AC 090901;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Shikimate kinase (EC 2.7.1.71) (SK).
GN AROK OR IL1743.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RA MEDLINE=21235186; PubMed=11337471;
RA Boctot A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -1- CATALYTIC ACTIVITY: ATP + shikimate = ADP + shikimate 3-phosphate.
CC -1- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE SHIKIMATE KINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
CC EMBL: ARO06404; AAK05841.1; -.
CC DR InterPro: IPR000623; Shik_kinase.
CC DR Pfam: PF01202; SKI; 1.
CC DR PRINTS: PR01100; SHIKIMKINASE.
CC DR PROSITE: PS01128; SHIKIMATE_KINASE; 1.
CC KW Aromatic amino acid biosynthesis; Transferase; Kinase; ATP-binding;
CC KW Complete proteome.
CC FT NP_BIND 7 14 ATP (POTENTIAL).
CC SO SEQUENCE 162 AA; 18383 MW; 1F7FDEFAF42AE9039 CRC64;

Query Match
Best Local Similarity 79.3%; Score 23; DB 1; Length 162;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
CC EMBL: AE000957; AAB89124.1; -.
CC DR HSSP: P11998; IRVV.
CC DR TIGR: AF2128; -.
CC DR InterPro: IPR002180; DMRL_synthase.
CC DR Pfam: PF00885; DMRL_synthase; 1.
CC KW Riboflavin biosynthesis; Transferase; Complete proteome.
CC SO SEQUENCE 143 AA; 15645 MW; BBA3141206905410 CRC64;

Query Match
Best Local Similarity 83.3%; Score 37;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MVXAEF 7
DB 8 MVXAEF 13

RESULT 14
AROK_LACLA STANDARD: PRT: 162 AA.
ID AROK_LACLA
AC 090901;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Shikimate kinase (EC 2.7.1.71) (SK).
GN AROK OR IL1743.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RA MEDLINE=21235186; PubMed=11337471;
RA Boctot A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -1- CATALYTIC ACTIVITY: ATP + shikimate = ADP + shikimate 3-phosphate.
CC -1- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE SHIKIMATE KINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
CC EMBL: ARO06404; AAK05841.1; -.
CC DR InterPro: IPR000623; Shik_kinase.
CC DR Pfam: PF01202; SKI; 1.
CC DR PRINTS: PR01100; SHIKIMKINASE.
CC DR PROSITE: PS01128; SHIKIMATE_KINASE; 1.
CC KW Aromatic amino acid biosynthesis; Transferase; Kinase; ATP-binding;
CC KW Complete proteome.
CC FT NP_BIND 7 14 ATP (POTENTIAL).
CC SO SEQUENCE 162 AA; 18383 MW; 1F7FDEFAF42AE9039 CRC64;

Query Match
Best Local Similarity 79.3%; Score 23; DB 1; Length 162;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

OY 2 MXVAEF 7  
1:111  
Db 37 MTIAEF 42

## RESULT 15

CYCR\_ERYSP STANDARD; PRT; 187 AA.

AC P26276; 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Photosynthetic reaction center cytochrome C subunit (Fragment).

CN Erythrobacter sp. (strain OCH 114).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;

OC Roseobacter.

OX NCBI\_TaxID=2434;

RN 11

RP MEDLINE=92157872; PubMed=1787796;

RA Liebetanz R., Hornberger U., Drews G.;

RT "Organization of the genes coding for the reaction-centre L and M

subunits and B870 antenna polypeptides alpha and beta from the

RT aerobic photosynthetic bacterium Erythrobacter species OCH114.";

RL Mol. Microbiol. 5:1459-1468(1991).

-1- FUNCTION: THE REACTION CENTER OF PURPLE BACTERIA CONTAIN A

TIGHTLY BOUND CYTOCHROME MOLECULE WHICH RE-REDUCES THE PHOTO

OXIDIZED PRIMARY ELECTRON DONOR.

-1- P1M: BINDS FOUR HEME GROUPS PER MOLECULE.

-1- SIMILARITY: HIGH, WITH OTHER PHOTOSYNTHETIC REACTION CENTER

CYTOCHROME C SUBUNITS.

-----

DR EMBL: X57597; CAA40820.1; -

DR PIR: S16314; S16314.

DR HSSP: P07173; 6PRC.

DR InterPro: IPR003158; CYTC\_RC.

DR InterPro: IPR000345; CYTC\_heme\_bind.

DR Pfam: PF02276; CytoC\_RC; 1.

DR PROSITE: PS00190; CYTOCHROME\_C; 2.

KW Electron transport; Photosynthesis; Reaction center; Heme;

Duplication. 127

FT BINDING 127

FT BINDING 130

FT METAL 131

FT BINDING 178

FT BINDING 181

FT METAL 182

FT NON\_TER 187

FT SEQUENCE 187 AA; 20681 MW; 8F06B6BEAC557D5 CRC64;

Query Match 79.3%; Score 23; DB 1; Length 187;

Best Local Similarity 83.3%; Pred. No. 48;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MXVAEF 7

1:111

Db 58 MHVAEF 63

Search completed: October 30, 2002, 12:27:50

Job time : 2.60197 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 2.18428 Seconds

(without alignments)  
554,401 Million cell updates/sec

Title: US-09-724-571-78

Perfect score: 29

Sequence: 1 VMXVAEF 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	96.6	776	10 Q941L7	Q941L7 populus tre
2	28	96.6	3414	12 Q91G40	Q91G40 langat viru
3	28	96.6	3414	12 Q91G39	Q91G39 langat viru
4	27	93.1	775	10 Q9FE38	Q9FE38 arabidopsi
5	27	93.1	802	10 Q9SUR2	Q9SUR2 arabidopsi
6	26	89.7	824	10 Q9ZP05	Q9ZP05 arabidopsi
7	26	89.7	821	2 Q9XW4	Q9XW4 porphyromon
8	25	86.2	201	16 Q34914	Q34914 bacillus su
9	25	86.2	1077	10 P92974	P92974 arabidopsi
10	24	82.8	195	16 Q97F65	Q97F65 clostridium
11	24	82.8	248	5 Q97008	Q97008 leishmania
12	24	82.8	249	16 Q9PKD5	Q9PKD5 chlamydia m
13	24	82.8	270	5 Q9NE74	Q9NE74 leishmania
14	24	82.8	317	5 Q9U375	Q9U375 caenorhabdi
15	24	82.8	335	16 Q97RC6	Q97RC6 streptococc
16	24	82.8	367	16 Q97KM0	Q97KM0 clostridium

17	24	82.8	378	10 Q9FNG6	Q9FNG6 arabidopsi
18	24	82.8	427	11 Q99L13	Q99L13 mus musculu
19	24	82.8	484	10 Q9C7R2	Q9C7R2 arabidopsi
20	24	82.8	582	17 Q58843	Q58843 pyrococcus
21	24	82.8	584	17 Q9UZM0	Q9UZM0 pyrococcus
22	24	82.8	738	5 P91063	P91063 caenorhabdi
23	24	82.8	817	5 Q9VZJ5	Q9VZJ5 drosophila
24	24	82.8	921	16 Q9K9Y0	Q9K9Y0 bacillus ha
25	24	82.8	1117	5 Q21477	Q21477 caenorhabdi
26	24	82.8	1122	5 Q61460	Q61460 caenorhabdi
27	24	82.8	1158	5 Q96554	Q96554 caenorhabdi
28	23	79.3	56	13 Q91995	Q91995 xenopus lae
29	23	79.3	57	11 Q9QW56	Q9QW56 mus sp. pro
30	23	79.3	57	13 Q08556	Q08556 phasiantdae
31	23	79.3	61	4 Q9P166	Q9P166 homo sapien
32	23	79.3	68	13 Q9Y121	Q9Y121 gallus gall
33	23	79.3	89	2 Q92GY2	Q92GY2 yersinia pe
34	23	79.3	105	2 Q91657	Q91657 salmonella
35	23	79.3	129	11 Q63000	Q63000 rattus norv
36	23	79.3	138	13 Q9PSE8	Q9PSE8 gallus gall
37	23	79.3	145	16 Q98BM6	Q98BM6 rhizobium l
38	23	79.3	154	10 Q9FHV3	Q9FHV3 arabidopsi
39	23	79.3	157	6 Q77807	Q77807 bos taurus
40	23	79.3	157	6 Q9N0K5	Q9N0K5 sus scrofa
41	23	79.3	157	6 Q28332	Q28332 callithrix
42	23	79.3	157	6 Q28111	Q28111 bos taurus
43	23	79.3	165	2 Q9F4D2	Q9F4D2 bacteroides
44	23	79.3	172	13 Q91176	Q91176 oryza lat
45	23	79.3	172	13 Q91177	Q91177 oryza lat

## ALIGNMENTS

RESULT 1  
Q941L7 PRELIMINARY; PRT; 776 AA.  
ID Q941L7  
AC Q941L7;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PUTATIVE HIGH-AFFINITY POTASSIUM UPTAKE TRANSPORTER.  
GN KUP1.  
OS Populus tremula x Populus tremuloides.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Malpighiales; Salicaceae; Populus.  
OX NCBI\_TaxID=47664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Langer K., Ache P., Fromm J., Hedrich R.;  
RT "PtkUP contributes to wood formation."  
RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ299422; CAC39168.1; -  
SQ SEQUENCE 776 AA; 87303 MW; BEC03D57ED0869BC CRC64;

Query Match 96.6%; Score 28; DB 10; Length 776;  
Best Local Similarity 85.7%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
DB 604 VMSVAEF 610  
RESULT 2  
Q91G40 PRELIMINARY; PRT; 3414 AA.  
AC Q91G40;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE POLYPROTEIN PRECURSOR.

```

OS   Langat virus (strain TP21).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
ON   NCBI_TaxID=31638;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TP21;
RX   MEDLINE=20192178; PubMed=10725214;
RA   Campbell M.S., Plenev A.G.;
RT   "Infectious cDNA clones of Langat tick-borne flavivirus that differ
RL   from their parent in peripheral neurovirulence.";
RL   Virology 269:225-237(2000).
DR   EMBL: AF253419; AAF75259.1; -.
DR   HSSP: P14336; 1SVB.
DR   MEROPS: S07.001; -.
DR   InterPro: IPR001410; DEAD.
DR   InterPro: IPR000336; Flavi_glycoprote.
DR   InterPro: IPR001850; Flavi_helicase.
DR   InterPro: IPR001157; Flavi_NSI.
DR   InterPro: IPR001528; Flavi_NS4B.
DR   InterPro: IPR000208; Flavi_NS5.
DR   InterPro: IPR001650; Helicase_C.
DR   Pfam: PF00869; Flavi_glycoprote.1.
DR   Pfam: PF00949; Flavi_helicase.1.
DR   Pfam: PF00948; Flavi_NSI.1.
DR   Pfam: PF01349; Flavi_NS4B.1.
DR   Pfam: PF00972; Flavi_NS5.1.
DR   ProDom: PD001496; Flavi_NSI.1.
DR   ProDom: PD001556; Flavi_glycoprote.1.
DR   SMART: SM00490; HELIC; 1.
KW   Signal.
FT   SIGNAL      97      116      POTENTIAL.
FT   CHAIN       1       96      CAPSID PROTEIN.
FT   CHAIN       117      280     MEMBRANE-ASSOCIATED GLYCOPROTEIN
FT   CHAIN       117      280     PRECURSOR.
FT   CHAIN       281      779     ENVELOPE MEMBRANE-ASSOCIATED
FT   CHAIN       281      779     GLYCOPROTEIN.
FT   CHAIN       780      1128    NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN       1129     1358    NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN       1359     1489    NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN       1490     2110    NONSTRUCTURAL PROTEIN NS3.
FT   CHAIN       2111     2259    NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN       2260     2511    NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN       2512     3414    NONSTRUCTURAL PROTEIN NS5.
FT   CHAIN       3414 AA; 378135 MW; 8DE86A46A08BF7E3 CRC64;
SQ   SEQUENCE

Query Match          96.6%; Score 28; DB 12; Length 3414;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY   1 VMXVAEF 7
      |||||
DB   929 VMTVAEF 935

RESULT 3
ID   Q9IG39      PRELIMINARY;      PRT; 3414 AA.
AC   Q9IG39;
DT   01-OCT-2000 (TREMBLrel. 15, Created)
DT   01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   POLYPROTEIN PRECURSOR.
OS   Langat virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
ON   NCBI_TaxID=11085;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATTENUATED STRAIN E5;
RX   MEDLINE=20192178; PubMed=10725214;
RA   Campbell M.S., Plenev A.G.;
RT   "Infectious cDNA clones of Langat tick-borne flavivirus that differ

```

```

RT   from their parent in peripheral neurovirulence.";
RL   Virology 269:225-237(2000).
DR   EMBL: AF253420; AAF75260.1; -.
DR   HSSP: P14336; 1SVB.
DR   MEROPS: S07.001; -.
DR   InterPro: IPR001410; DEAD.
DR   InterPro: IPR000336; Flavi_glycoprote.
DR   InterPro: IPR001850; Flavi_helicase.
DR   InterPro: IPR001157; Flavi_NSI.
DR   InterPro: IPR001528; Flavi_NS4B.
DR   InterPro: IPR000208; Flavi_NS5.
DR   InterPro: IPR001650; Helicase_C.
DR   Pfam: PF00869; Flavi_glycoprote.1.
DR   Pfam: PF00949; Flavi_helicase.1.
DR   Pfam: PF00948; Flavi_NSI.1.
DR   Pfam: PF01349; Flavi_NS4B.1.
DR   Pfam: PF00972; Flavi_NS5.1.
DR   ProDom: PD001496; Flavi_NSI.1.
DR   ProDom: PD001556; Flavi_glycoprote.1.
DR   SMART: SM00490; HELIC; 1.
KW   Signal.
FT   SIGNAL      97      116      POTENTIAL.
FT   CHAIN       1       96      CAPSID PROTEIN.
FT   CHAIN       117      280     MEMBRANE-ASSOCIATED GLYCOPROTEIN
FT   CHAIN       117      280     PRECURSOR.
FT   CHAIN       281      779     ENVELOPE MEMBRANE-ASSOCIATED
FT   CHAIN       281      779     GLYCOPROTEIN.
FT   CHAIN       780      1128    NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN       1129     1358    NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN       1359     1489    NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN       1490     2110    NONSTRUCTURAL PROTEIN NS3.
FT   CHAIN       2111     2259    NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN       2260     2511    NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN       2512     3414    NONSTRUCTURAL PROTEIN NS5.
FT   CHAIN       3414 AA; 378014 MW; BF8F8ACEB96D534 CRC64;
SQ   SEQUENCE

Query Match          96.6%; Score 28; DB 12; Length 3414;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY   1 VMXVAEF 7
      |||||
DB   929 VMTVAEF 935

RESULT 4
ID   Q9FE38      PRELIMINARY;      PRT; 775 AA.
AC   Q9FE38;
DT   01-MAR-2001 (TREMBLrel. 16, Created)
DT   01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   TINY ROOT HAIR 1 PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsids.
ON   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rigas S., Debrosses G., Haralampidis K., Vicente-Aguilo F.,
RA   Feldmann K.A., Grabov A., Dolan L., Hatzopoulos P.;
RT   "A Potassium Transporter Required for Tip Growth in Arabidopsis.";
RL   Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR   EMBL: AJ296156; CAC16138.1; -.
DR   EMBL: AJ296155; CAC16137.1; -.
DR   InterPro: IPR003855; K+-transprt.
DR   Pfam: PF02705; K.trans.1.
SQ   SEQUENCE 775 AA; 86842 MW; B0C55068B48E8180 CRC64;

Query Match          93.1%; Score 27; DB 10; Length 775;
Best Local Similarity 71.4%; Pred. No. 99;

```



Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:||||  
Db 602 VMSIAEF 608

## RESULT 5

O9SUR2 PRELIMINARY; PRT; 802 AA.

AC O9SUR2;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
GN PUTATIVE POTASSIUM TRANSPORT PROTEIN.  
GN F9D16.110 OR A74G23640.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RA SEQUENCE FROM N.A.  
RA Bevan M., Medler H., Medler E., Wambutt R., Hohnsels J., Mewes H.W.,  
RA Mayer K.F.X., Scheller C.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RA SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RA SEQUENCE FROM N.A.  
RA Medler H., Medler E., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RA SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL055394; CAA23030.1; -  
DR EMBL: AL161559; CAB79319.1; -  
DR InterPro: IPR003855; K+-transprtr.  
DR Pfam: PF02705; K-trans; 1.  
SQ SEQUENCE 802 AA; 89817 MW; 617AF5F76B99B60 CRC64;

Query Match  
Best Local Similarity 93.1%; Score 27; DB 10; Length 802;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:||||  
Db 629 VMSIAEF 635

## RESULT 6

O9ZPQ5 PRELIMINARY; PRT; 324 AA.

AC O9ZPQ5;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
GN PUTATIVE STEROID SULFOTRANSFERASE.  
GN AT2G03770.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RA SEQUENCE FROM N.A.  
RA STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana";  
RL Nature 402:761-768(1999).

RN [2]  
RA SEQUENCE FROM N.A.  
RA STRAIN=CV. COLUMBIA;  
RA Lin X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC006836; AAD20079.1; -  
DR HSSP: P50224; 1C3M.  
DR InterPro: IPR000863; Sulfotransferase.  
DR Pfam: PF00685; Sulfotransfer; 1.  
DR ProDom: PD001218; Sulfotransfer; 1.  
KW Transferase.  
SQ SEQUENCE 324 AA; 37717 MW; B488C799AB13E1D3 CRC64;

Query Match  
Best Local Similarity 89.7%; Score 26; DB 10; Length 324;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:||||  
Db 233 VMRIAEF 239

## RESULT 7

O9XBW4 PRELIMINARY; PRT; 821 AA.

AC O9XBW4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
GN IMMUNOREACTIVE 92 KDA ANTIGEN PG21.  
OS Porphyromonas gingivalis (Bacteroides gingivalis).  
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;  
OC Porphyromonas.  
OX NCBI\_TaxID=837;  
RN [1]  
RA SEQUENCE FROM N.A.  
RA STRAIN=W50;  
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,  
RA Hocking D., Webb E.;  
RT "Porphyromonas gingivalis polypeptides and nucleic acids";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF153768; AAD38980.1; -  
DR MEROPS: M14.0PB; -  
DR InterPro: IPR000601; PKD\_domain.  
DR InterPro: IPR000834; Zn\_carpodpept.  
DR Pfam: PF00801; PKD; 3.  
DR Pfam: PF00246; Zn\_carpodpept; 1.  
DR PRINTS: PR00765; CRBOXYPTASEA.  
DR SMART: SM00089; PKD; 3.  
DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
DR PROSITE: PSS0093; PKD; 3.  
SQ SEQUENCE 821 AA; 91517 MW; 250843B2C9C83E2 CRC64;

Query Match  
Best Local Similarity 89.7%; Score 26; DB 2; Length 821;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXVAEF 7  
||:||||  
Db 565 VMPVAEF 571

RESULT 8  
O34914 PRELIMINARY; PRT; 201 AA.



```

OC Bacteria: Firmicutes: Bacillus/Clostridium group: Clostridiaceae;
OC Clostridium;
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359326; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007786; AAK80830.1; -.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001267; Thymd_kin_cell.
DR Pfam: PF00265; TK; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00603; TK_CELULAR_TYPE; UNKNOWN_1.
KW Kinase; Complete proteome.
SQ SEQUENCE 195 AA; 22303 MW; 6EDB5198B203BA CRC64;

Query Match
Best Local Similarity 82.8%; Score 24; DB 16; Length 195;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXYAEF 7
Db 131 LMAIAEF 137

RESULT 11
O97008 PRELIMINARY; PRT; 248 AA.
AC O97008;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 26.4 KDA PROTEIN.
GN L7610.05.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Ralndream M.A., Barrell B.G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL139794; CAC22635.1; -.
KW Hypothetical protein.
SQ SEQUENCE 248 AA; 26365 MW; AB0440E1B3F69B5 CRC64;

Query Match
Best Local Similarity 82.8%; Score 24; DB 5; Length 248;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXYAEF 7
Db 196 LMRVAEF 202

RESULT 12
O9PKD5

```

```

ID O9PKD5 PRELIMINARY; PRT; 249 AA.
AC O9PKD5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SERINE/THREONINE PROTEIN PHOSPHATASE, PUTATIVE.
GN TC0330.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MORN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002321; AAF39372.1; -.
DR TIGR: TC0530; -.
DR InterPro: IPR001932; PP2C_domain.
DR InterPro: IPR003643; Putat_Ppase.
DR ProDom: PD006823; Putat_Ppase.
DR SMART: SM00352; PP2C; 1.
DR SMART: SM00331; PP2C_SIG; 1.
KW Complete proteome.
SQ SEQUENCE 249 AA; 28192 MW; 6BD227BD3281164F CRC64;

Query Match
Best Local Similarity 82.8%; Score 24; DB 16; Length 249;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXYAEF 7
Db 1 MMYVAEF 7

RESULT 13
O9NE74 PRELIMINARY; PRT; 270 AA.
ID O9NE74;
AC O9NE74;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 29.3 KDA PROTEIN.
GN L5075.02.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Beck A., Klages S., Reinhardt R., Ivens A.C., Quail M.,
RA Ralndream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL163552; CAB86964.1; -.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 29335 MW; 5B91DA37A16FC3CB CRC64;

Query Match
Best Local Similarity 82.8%; Score 24; DB 5; Length 270;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VMXYAEF 7
Db 196 LMRVAEF 202

RESULT 12
O9PKD5

```

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
11:111  
Db 148 VMLAEF 154

## RESULT 14

Q9U375 PRELIMINARY; PRT; 317 AA.  
AC Q9U375;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)  
DE T08D2.2 PROTEIN.  
GN T08D2.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McMurtry A.A.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL; Z92839; CAB07417.2; -.  
DR InterPro: IPR001810; F-box.  
DR Pfam: PF00646; F-box; 1.  
SQ SEQUENCE 317 AA; 36591 MW; 145E8B02B75AE42 CRC64;

Query Match 82.8%; Score 24; DB 5; Length 317;  
Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 MXVAEF 7  
11111  
Db 223 MSVAEF 228

## RESULT 15

Q97RC6 PRELIMINARY; PRT; 335 AA.  
AC Q97RC6;  
DT 01-OCT-2001 (TReMBLrel. 18, Created)  
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE 6-PHOSPHOFRUCTOKINASE.  
GN SP0896.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., C.L.,  
RA Holtzapple E., Khouli H., Wolf A.M., Uterback T.R., Hansen C.L.,  
RA McDonald L.A., Peidilyum T.V., Angiolli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
pneumoniae."  
RL Science 293:498-506(2001).

DR EMBL; AE007395; AAK75023.1; -.  
DR TIGR; SP0896; -.  
DR InterPro: IPR000023; Phosphofructokinase.  
DR Pfam: PF00365; PFK; 1.  
DR PRINTS; PR00476; PHERCTKINASE.  
DR ProDom: PD000707; Phosphofructokinase; 1.  
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.  
KW Kinase; Complete proteome.  
SQ SEQUENCE 335 AA; 35174 MW; 2DFBACC58CAF66 CRC64;

Query Match 82.8%; Score 24; DB 16; Length 335;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VMXVAEF 7  
11111  
Db 224 VMSAEF 230

Search completed: October 30, 2002, 12:30:07  
Job time : 6.18428 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 2.80835 Seconds  
(without alignments)  
554.401 Million cell updates/sec

Title: US-09-724-571-91

Perfect score: 42

Sequence: 1 SEVNILAEF 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriap:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	83.3	649	5 Q9V7X7	Q9V7X7 drosophila
2	35	83.3	948	12 Q9QW73	Q9QW73 avian adeno
3	34	81.0	529	12 Q39281	Q39281 equine hepr
4	33	78.6	668	16 Q97KW1	Q97KW1 clostridium
5	33	78.6	837	4 Q9UH51	Q9UH51 homo sapien
6	32	76.2	132	16 Q9WZB6	Q9WZB6 thermotoga
7	32	76.2	279	10 Q65159	Q65159 phaseolus v
8	32	76.2	422	10 Q49336	Q49336 spinacia ol
9	32	76.2	467	10 Q9SSE8	Q9SSE8 phaseolus a
10	32	76.2	467	10 Q41688	Q41688 vigna radia
11	32	76.2	472	10 Q24544	Q24544 phaseolus a
12	32	76.2	472	10 Q41687	Q41687 vigna radia
13	32	76.2	476	16 Q92812	Q92812 rhizobium m
14	32	76.2	942	12 Q96585	Q96585 avian adeno
15	31	73.8	151	10 Q9FL93	Q9FL93 arabidopsis
16	31	73.8	167	2 P77934	P77934 photobacter

17	31	73.8	175	10 Q93VA3	Q93VA3 arabidopsis
18	31	73.8	210	10 Q9FKX3	Q9FKX3 arabidopsis
19	31	73.8	225	12 Q84614	Q84614 parametium
20	31	73.8	246	13 Q9W6E5	Q9W6E5 gallus gall
21	31	73.8	295	10 Q9SHF7	Q9SHF7 arabidopsis
22	31	73.8	384	4 Q9EQN9	Q9EQN9 homo sapien
23	31	73.8	389	13 Q9W623	Q9W623 xenopus lae
24	31	73.8	393	11 Q91XQ1	Q91XQ1 rattus norv
25	31	73.8	394	16 Q9A4F6	Q9A4F6 caulobacter
26	31	73.8	397	2 Q9S523	Q9S523 escherichia
27	31	73.8	411	2 Q68771	Q68771 yersinia pe
28	31	73.8	411	2 Q935E6	Q935E6 salmonella
29	31	73.8	435	3 Q74763	Q74763 schizosacch
30	31	73.8	442	13 Q9W622	Q9W622 xenopus lae
31	31	73.8	443	11 Q921N9	Q921N9 mus musculu
32	31	73.8	443	11 Q91218	Q91218 mus musculu
33	31	73.8	906	5 Q17281	Q17281 botryllus s
34	31	73.8	1357	12 Q89328	Q89328 rice ragged
35	31	73.8	2270	12 Q9JFN3	Q9JFN3 tupala para
36	31	73.8	3169	2 Q92GR4	Q92GR4 escherichia
37	31	73.8	3169	2 Q92GR4	Q92GR4 escherichia
38	30.5	72.6	557	13 Q90252	Q90252 brachydanio
39	30	71.4	118	16 Q988K6	Q988K6 rhizobium l
40	30	71.4	133	16 Q9CRU2	Q9CRU2 lactococcus
41	30	71.4	218	8 Q98455	Q98455 spirogyra m
42	30	71.4	293	5 Q9VJZ7	Q9VJZ7 drosophila
43	30	71.4	298	11 Q9EQ06	Q9EQ06 mus musculu
44	30	71.4	300	4 Q9UR04	Q9UR04 homo sapien
45	30	71.4	300	4 Q96HF6	Q96HF6 homo sapien

## ALIGNMENTS

RESULT 1	ID	Q9V7X7	PRELIMINARY:	PRT:	649 AA.
Q9V7X7	AC	Q9V7X7			
DT	01-MAY-2000	(TREMblrel. 13, Created)			
DT	01-MAY-2000	(TREMblrel. 13, Last sequence update)			
DT	01-DEC-2001	(TREMblrel. 19, Last annotation update)			
DE	CG8961	PROTEIN (LD32631P).			
GN	TEF OR CG8961.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BERKELEY.				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,				
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abriil J.F., Agbayani A., An H.-C., Andrews-Plankk C., Baldwin D.,				
RA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,				
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,				
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,				
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hostli M., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,				

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshireti A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Plattman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Turner C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhang M., Zhou X., Zhu S., Zhu X.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster.",  
RL Science 287:2185-2195(2000).  
[2]  
RV SEQUENCE FROM N.A.  
RP STRAIN=Y, CN BW SP;  
RC Stapleton M., Brokstein P., Hong L., Abghayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.,  
RA Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003804; AAF57911.1; -  
DR EMBL; AY051818; AAK93242.1; -  
DR Flybase; FBgn0043472; tef.  
DR InterPro; IPR000822; znf-C2H2.  
DR Pfam; PF00096; znf-C2H2; 3.  
DR SMART; SM00355; znf-C2H2; 3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 1.  
KW DNA-binding; Metal-binding; Zinc-finger.  
SQ SEQUENCE 649 AA; 73357 MW; F96518EB37D60E68 CRC64;

Query Match 83.3%; Score 35; DB 5; Length 649;  
Best Local Similarity 87.5%; Pred. No. 27;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLAIEF 8  
DB 131 SEINLAIEF 138

RESULT 2  
O9QM73 PRELIMINARY; PRT; 948 AA.  
AC O9QM73;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE P1.  
OS Avian adenovirus type 8 (strain ATCC A-2a) (Fowl adenovirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.  
OX NCBI\_TaxID=66295;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=20318671; PubMed=10859390;  
RX Ojikic D., Nagy E.;  
RT "The complete nucleotide sequence of fowl adenovirus type 8.";  
J Gen. Virol. 81:1833-1837(2000).  
[2]  
RV SEQUENCE FROM N.A.  
RP Ojikic D., Nagy E.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF083975; AAD50344.2; -  
DR HSSP; P03277; IDHx.  
DR InterPro; IPR000736; Adeno\_hexon.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF01065; Adeno\_hexon; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.

DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
SQ SEQUENCE 948 AA; 107604 MW; 67146E79F31446B8 CRC64;  
Query Match 83.3%; Score 35; DB 12; Length 948;  
Best Local Similarity 77.8%; Pred. No. 41;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAIEF 9  
DB 587 SEVNLAIEF 595

RESULT 3  
O39281 PRELIMINARY; PRT; 529 AA.  
AC O39281;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE COUNTERPART OF HSV-1 GENE UL21 AND VZV GENE 38.  
GN 40.  
OS Equine herpesvirus 4.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=10331;  
RN [1]  
RP SEQUENCE OF 489-529 FROM N.A.  
RP STRAIN=NS80567;  
RC MEDLINE=90362066; PubMed=2167933;  
RX Nicolson L., Cullinane A.A., Onions D.E.;  
RT "The nucleotide sequence of an equine herpesvirus 4 gene homologue of the herpes simplex virus 1 glycoprotein H gene";  
J Gen. Virol. 71:1793-1800(1990).  
[2]  
RV SEQUENCE FROM N.A.  
RP STRAIN=NS80567;  
RC MEDLINE=98266497; PubMed=9603335;  
RX Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;  
RT "The DNA sequence of equine herpesvirus-4.";  
J Gen. Virol. 79:1197-1203(1998).  
[3]  
RV SEQUENCE FROM N.A.  
RP STRAIN=NS80567;  
RC Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF030027; AAC59558.1; -  
SQ SEQUENCE 529 AA; 58150 MW; E2B737B9D489413 CRC64;

Query Match 81.0%; Score 34; DB 12; Length 529;  
Best Local Similarity 87.5%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVNLAIEF 9  
DB 423 EVNLAIEF 430

RESULT 4  
O97KW1 PRELIMINARY; PRT; 668 AA.  
AC O97KW1;  
DT 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE METHYL-ACCEPTING CHEMOTAXIS PROTEIN.  
GN CAC0805.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RX MEDLINE-21359325; PubMed-11466286;  
 RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium *Clostridium acetobutylicum*.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007596; AAK78781.1; -.  
 DR InterPro: IPR004010; Cache.  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMF.  
 DR InterPro: IPR003122; TsrH.  
 DR Pfam; PF02743; Cache; 1.  
 DR Pfam; PF00672; HAMF; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR SMART; SM00304; HAMF; 1.  
 DR SMART; SM00283; MA; 1.  
 DR SMART; SM00319; TsrH; 1.  
 KM Complete proteome.  
 SO SEQUENCE 668 AA; 73157 MW; F897FD68E105D6FF CRC64;

Query Match 78.6%; Score 33; DB 16; Length 668;  
 Best Local Similarity 75.0%; Pred. No. 80;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLAE 8  
 DB 71 SEINLMAE 78

RESULT 5  
 ID O9UH51 PRELIMINARY; PRT; 837 AA.  
 AC O9UH51;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE LOW DENSITY LIPOPROTEIN RECEPTOR (FRAGMENT).  
 GN LDLR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Van Leuven F., Thiry E., Stas L., Lambrechts M., Boon T.,  
 RA Bruynseels K., Descamps O.;  
 RT "Sequencing the coding exons for mutational analysis of the LRP1 and  
 RT LDLR genes on individual DNA samples.";  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF217403; AAF24515.1; -.  
 DR HSSP; P01130; LDL.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR000033; Ldl\_receptor\_rep.  
 DR InterPro: IPR002172; Ldl\_recept\_A.  
 DR Pfam; PF00008; EGF; 3.  
 DR Pfam; PF00057; Ldl\_recept\_a; 7.  
 DR Pfam; PF00058; Ldl\_recept\_b; 5.  
 DR PRINTS; PR00261; LDURECEPT.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF\_like; 2.  
 DR SMART; SM00192; LDla; 7.  
 DR SMART; SM00135; LY; 5.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_2.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01209; LDLRA\_1; 7.  
 DR PROSITE; PS00068; LDLRA\_2; 7.  
 KM Calcium-binding; EGF-like domain; Glycoprotein; Lipoprotein; Receptor;  
 Repeat.

FT NON\_TER 1 1  
 SQ SEQUENCE 837 AA; 92914 MW; 22AFB9382D7CE656 CRC64;

Query Match 78.6%; Score 33; DB 4; Length 837;  
 Best Local Similarity 87.5%; Pred. No. 1e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLAE 8  
 DB 614 SDVNLMAE 621

RESULT 6  
 ID O9WZB6 PRELIMINARY; PRT; 132 AA.  
 AC O9WZB6;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN.  
 GN TM0648.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE-99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smit H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of *Thermotoga maritima*.";  
 RL Nature 399:323-329(1999).  
 DR EMBL; AE001738; AAD35732.1; -.  
 DR TIGR; TM0648; -.  
 DR InterPro: IPR002934; NTP\_transf.  
 DR Pfam; PF01909; NTP\_Transf\_2; 1.  
 KM Complete proteome.  
 SO SEQUENCE 132 AA; 15264 MW; 4E973A6234276462 CRC64;

Query Match 76.2%; Score 32; DB 16; Length 132;  
 Best Local Similarity 66.7%; Pred. No. 25;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 DB 86 SEINLSEF 94

RESULT 7  
 ID O65159 PRELIMINARY; PRT; 279 AA.  
 AC O65159;  
 DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE 1-AMINOOCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE (FRAGMENT).  
 GN ACS1.  
 OS Phaseolus vulgaris (Kidney bean) (French bean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
 OX NCBI\_TaxID=3885;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. TAYLOR;  
 RA Pidgeon C.M., Reid D.M.;  
 RT "The role of light in controlling ACC synthase activity and gene  
 expression in *Phaseolus vulgaris* cv. Taylor.";

Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE TWO ISOZYMES: A CYTOPLASMIC ONE AND A MITOCHONDRIAL ONE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC EMBL: AF053355; AAC08356.1; -.  
 DR HSSP: P37821; 1B8G.  
 DR InterPro: IPR001176; ACC\_synthase.  
 DR InterPro: IPR001511; Aminotran\_1.  
 DR Pfam: PF00155; aminotran\_1\_2; 1.  
 DR PRINTS: PR00753; ACCSYNTHASE.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 DR Pyridoxal phosphate.. 1  
 FT NON\_TER 279 279  
 FT SEQUENCE 279 AA; 31069 MW; 50AE1FF551CFA7 CRC64;  
 SO  
 Query Match 76.2%; Score 32; DB 10; Length 279;  
 Best Local Similarity 66.7%; Pred. No. 54;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLLAE 9  
 DB 129 SEVNLVDF 137

RESULT 8  
 O49936 PRELIMINARY; PRT; 422 AA.  
 ID 049936;  
 AC 049936;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE BASIC LEUCINE ZIPPER PROTEIN.  
 GN BZIP.  
 OS Spinacia oleracea (Spinach).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OC NCBI\_TaxID=3562;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bolle C., Lueberstedt T., Herranen M., Herrmann R., Oelmueller R.,  
 RT "Molecular characterization of the spinach G-box binding protein family.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.  
 CC EMBL: AJ223624; CA011499.1; -.  
 DR HSSP: P03069; 1DGC.  
 DR InterPro: IPR001871; bZIP.  
 DR Pfam: PF00170; bZIP; 1.  
 DR SMART: SM00338; BRIZ; 1.  
 DR PROSITE: PS00036; BZIP\_BASIC; 1.  
 DR DNA-binding; Nuclear protein.  
 KW SEQUENCE 422 AA; 44924 MW; F9713C3BCA65D2F0 CRC64;  
 SO  
 Query Match 76.2%; Score 32; DB 10; Length 422;  
 Best Local Similarity 87.5%; Pred. No. 83;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLLAE 8  
 DB 324 SEVNLVDF 331

RESULT 9  
 O9SEB8 PRELIMINARY; PRT; 467 AA.  
 ID 09SEB8;  
 AC 09SEB8;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE.  
 GN ACS7.  
 OS Phaseolus aureus (Mung bean) (Vigna radiata).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
 OC NCBI\_TaxID=3916;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-20074367; PubMed-10608655;  
 RX Yi H.C., Joo S., Nam K.H., Lee J.S., Kang B.G., Kim W.T.;  
 RT "Auxin and brassinosteroid differentially regulate the expression of three members of the 1-aminocyclopropene-1-carboxylate synthase gene family in mung bean (Vigna radiata L.).";  
 RT Plant Mol. Biol. 41:443-454(1999).  
 RL  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE TWO ISOZYMES: A CYTOPLASMIC ONE AND A MITOCHONDRIAL ONE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC EMBL: AF151961; AAD41083.1; -.  
 DR HSSP: P37821; 1B8G.  
 DR InterPro: IPR001176; ACC\_synthase.  
 DR InterPro: IPR001511; Aminotran\_1.  
 DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00155; aminotran\_1\_2; 1.  
 DR PRINTS: PR00753; ACCSYNTHASE.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
 DR Pyridoxal phosphate.. 1  
 KW SEQUENCE 467 AA; 53044 MW; 121423C4ABACB079 CRC64;  
 SO  
 Query Match 76.2%; Score 32; DB 10; Length 467;  
 Best Local Similarity 66.7%; Pred. No. 93;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLLAE 9  
 DB 211 SEVNLVDF 219

RESULT 10  
 Q41688 PRELIMINARY; PRT; 467 AA.  
 ID Q41688;  
 AC Q41688;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE.  
 GN VRACS7.  
 OS Vigna radiata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
 OC NCBI\_TaxID=157791;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE-ETIOLOGICAL HYPOCOTYL;  
 RA Kim W.;  
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE TWO ISOZYMES: A CYTOPLASMIC ONE AND A MITOCHONDRIAL ONE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC EMBL: U34987; AAT8274.1; -.  
 DR HSSP: P37821; 1B8G.  
 DR InterPro: IPR001176; ACC\_synthase.  
 DR InterPro: IPR001511; Aminotran\_1.



DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00155; aminotran\_1.2; 1.  
 DR PRINTS: PR00753; ACCSYNTHASE.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
 KW Pyridoxal phosphate.  
 SO SEQUENCE 467 AA; 53058 MW; 16710824A0376A79 CRC64;

Query Match 76.2%; Score 32; DB 10; Length 467;  
 Best Local Similarity 66.7%; Pred. No. 93;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 ||:||||:|  
 Db 211 SELNLVDF 219

## RESULT 11

ID 024544 PRELIMINARY; PRT; 472 AA.  
 AC 024544;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE..  
 GN VR-ACS6.  
 OS Phaseolus aureus (Mung bean) (Vigna radiata).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
 OX NCBI\_TaxID=3916;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ETIOLATED HYPOCOTYL;  
 RX MEDLINE=97294927; PubMed=9150600;  
 RA Yoon I., Mori H., Kim J., Kang B., Imaseki H.;  
 RT "VR-ACS6 is an auxin-inducible 1-aminocyclopropene-1-carboxylate  
 RT synthase gene in mungbean (Vigna radiata).";  
 RL Plant Cell Physiol. 38:217-224(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HYPOCOTYL;  
 RA Yoon I.S., Park D.H., Mori H., Imaseki H., Kang B.G.;  
 RT "Characterization of an auxin-inducible ACC synthase gene, VR-ACS6 in  
 RT mungbean and expression of its promoter in transgenic tobacco.";  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CORCTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE TWO ISOZYMES: A CYTOPLASMIC  
 CC ONE AND A MITOCHONDRIAL ONE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC EMBL: AB000679; BAA19161.1; -;  
 DR EMBL: AB018355; BAA33859.1; -;  
 DR HSSP: P37821; 1B8G.  
 DR InterPro: IPR001176; ACC\_synthase.  
 DR InterPro: IPR001511; Aminotran\_1.  
 DR Pfam: PF00155; aminotran\_1.2; 1.  
 DR PRINTS: PR00753; ACCSYNTHASE.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 KW Pyridoxal phosphate.  
 SO SEQUENCE 472 AA; 53572 MW; 1404AD4B3BD0421 CRC64;

Query Match 76.2%; Score 32; DB 10; Length 472;  
 Best Local Similarity 66.7%; Pred. No. 94;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 ||:||||:|  
 Db 211 SELNLVDF 219

## RESULT 12

ID 041687 PRELIMINARY; PRT; 472 AA.  
 AC 041687;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE.  
 GN VRACS6.  
 OS Vigna radiata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
 OX NCBI\_TaxID=157791;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ETIOLATED HYPOCOTYLS;  
 RA Kim W.;  
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE TWO ISOZYMES: A CYTOPLASMIC  
 CC ONE AND A MITOCHONDRIAL ONE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC EMBL: U34986; AAA78273.1; -;  
 DR HSSP: P37821; 1B8G.  
 DR InterPro: IPR001176; ACC\_synthase.  
 DR InterPro: IPR001511; Aminotran\_1.  
 DR Pfam: PF00155; aminotran\_1.2; 1.  
 DR PRINTS: PR00753; ACCSYNTHASE.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 DR PROSITE: PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 KW Pyridoxal phosphate.  
 SO SEQUENCE 472 AA; 53518 MW; 0C5F3E22AB0E8EAF CRC64;

Query Match 76.2%; Score 32; DB 10; Length 472;  
 Best Local Similarity 66.7%; Pred. No. 94;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 ||:||||:|  
 Db 211 SELNLVDF 219

## RESULT 13

ID 092ZL2 PRELIMINARY; PRT; 476 AA.  
 AC 092ZL2;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NIFE OXIDOREDUCTASE.  
 GN NIFE.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC plasmid pSyma (megaplasmid 1).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SRRATN=1021;  
 RC MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Bariloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,  
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kalmann S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Voh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSyma megaplasmid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 DR EMBL: AE007236; AAK65110.1; -;  
 KW plasmid; Complete proteome.  
 SO SEQUENCE 476 AA; 52358 MW; 83ECE94B4B14C713 CRC64;

Query Match 76.2%; Score 32; DB 16; Length 476;  
 Best Local Similarity 55.6%; Pred. No. 95;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLLAEF 9  
 ID 195 SDINIGEF 203

## RESULT 14

Q96585 PRELIMINARY; PRT; 942 AA.  
 AC Q96585;  
 DT 01-FEB-1997 (TREMblrel. 02, Created)  
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE VIRAL STRUCTURAL PROTEIN.  
 GN HEXON.  
 OS Avian adenovirus gall (strain Phelps) (Powl adenovirus 1) (CELO).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.  
 OX NCBI\_TaxID=10553;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FAV1 (CELO);  
 RX MEDLINE=97049067; PubMed=8893797;  
 RA Akopian T.A., Doronin K.K., Karpov V.A., Naroditsky B.S.;  
 RT "Sequence of the avian adenovirus FAV1 (CELO) DNA encoding the hexon-  
 associated protein pIV and hexon."  
 RL Arch. Virol. 141:1759-1765(1996).  
 DR EMBL; Z67970; CAA91908.1; -.  
 DR HSP; P03277; IDHX.  
 DR InterPro; IPR00736; Adeno\_hexon.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF01065; Adeno\_hexon; 1.  
 DR ProDom; PD002815; Adeno\_hexon; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 SQ SEQUENCE 942 AA; 106695 MW; 903732D3F5E6DC6F CRC64;

Query Match 76.2%; Score 32; DB 12; Length 942;  
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLLAEF 9  
 ID 581 TEVNLAMNF 589

## RESULT 15

Q9FL93 PRELIMINARY; PRT; 151 AA.  
 AC Q9FL93;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE GENOMIC DNA, CHROMOSOME 5, TAC CLONE:K21C13.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=98344145; PubMed=9679202;  
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence  
 features of the regions of 1,381,565 bp covered by twenty one  
 physically assigned pl and TAC clones."  
 RL DNA Res. 5:131-145(1998).  
 DR EMBL; AB010693; BAB10887.1; -.  
 DR HSP; P08197; ICYU.  
 DR InterPro; IPR000345; CytC\_heme\_bind.

DR InterPro; IPR003088; Cyt\_C1.  
 DR Pfam; PF00034; CYTOCHROME\_C\_1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 SQ SEQUENCE 151 AA; 17031 MW; 2C7F3A4E4ADEDE2 CRC64;

Query Match 73.8%; Score 31; DB 10; Length 151;  
 Best Local Similarity 75.0%; Pred. No. 48;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVNLLAEF 9  
 ID 129 EIKLLAEF 136

Search completed: October 30, 2002, 12:30:13  
 Job time : 5.80835 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:10 ; Search time 0.773956 Seconds  
(without alignments)  
450.253 Million cell updates/sec

Title: US-09-724-571-91

Perfect score: 42

Sequence: 1 SEVNLLAEF 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	78.6	860	1	LDLR_HUMAN
2	32	76.2	476	1	NIFE_RHIME
3	32	76.2	942	1	HEX_ADECI
4	31	73.8	181	1	ORX_VIBCH
5	31	73.8	443	1	NEK2_MOUSE
6	31	73.8	445	1	NEK2_HUMAN
7	30	71.4	103	1	Y041_AQUAE
8	30	71.4	328	1	Y778_AQUAE
9	30	71.4	581	1	Y015_SCHPO
10	30	71.4	769	1	MEC4_CAEEL
11	30	71.4	769	1	MEC4_CAEEL
12	30	71.4	854	1	LDLR_CRIGR
13	30	71.4	864	1	LDLR_MOUSE
14	30	71.4	890	1	IF2_CHLPN
15	29	69.0	135	1	Y549_RICPR
16	29	69.0	230	1	R33_PORPU
17	29	69.0	338	1	YABT_BACSV
18	29	69.0	405	1	CPXK_SACER
19	29	69.0	460	1	UDPE_GYLO
20	29	69.0	613	1	ENP4_MOUSE
21	29	69.0	616	1	ENP4_HUMAN
22	29	69.0	1666	1	CO3_CAVPO
23	29	69.0	5430	1	ACE7_HUMAN
24	28	66.7	237	1	ORN_HUMAN
25	28	66.7	250	1	Y006_ANASP
26	28	66.7	302	1	BIRA_HAEIN
27	28	66.7	306	1	BUB2_YEAST
28	28	66.7	338	1	CYSP_ECOLI
29	28	66.7	338	1	CYSP_ECOLI
30	28	66.7	339	1	CYSP_SALTY
31	28	66.7	391	1	Y773_CAEEL
32	28	66.7	394	1	ACKA_LACSK
33	28	66.7	419	1	P47K_PSECL

34	28	66.7	437	1	V014_FOWPY	Q91517 fow1pox vir
35	28	66.7	457	1	ERF1_GIALA	O9ncp1 giardia lam
36	28	66.7	496	1	NIFE_RHIO	O9napi rhizobium l
37	28	66.7	496	1	NIFE_RHISN	P55673 rhizobium s
38	28	66.7	530	1	UL21_HSVB	P28972 equine herp
39	28	66.7	624	1	SERA_ARATH	O04130 arabidopsis
40	28	66.7	778	1	DEG1_CAEEL	P24585 caenorhabdi
41	28	66.7	804	1	SCY1_YEAST	P53009 saccharomyc
42	28	66.7	865	1	TOP1_ECOLI	P06612 escherichia
43	28	66.7	868	1	TOP1_PASMU	O9cnc30 pasteurilla
44	28	66.7	874	1	STA_HAEIN	P43815 haemophilus
45	28	66.7	1502	1	MT70_YEAST	P38181 saccharomyc

## ALIGNMENTS

RESULT 1

ID	LDLR_HUMAN	STANDARD	PRT	860 AA.
AC	P01130:			
DT	21-JUL-1986 (rel. 01, Created)			
DT	21-JUL-1986 (rel. 01, Last sequence update)			
DT	16-OCT-2001 (rel. 40, Last annotation update)			
DE	Low-density lipoprotein receptor precursor (LDL receptor).			
GN	LDLR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85024898; PubMed=6091915;			
RA	Yamamoto T., Davis C.G., Brown M.S., Schneider W.J., Casey M.L.,			
RA	Goldstein J.L., Russell D.W.;			
RT	"The human LDL receptor: a cysteine-rich protein with multiple Alu			
RT	sequences in its mRNA.";			
RL	Cell 39:27-38(1984).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85218750; PubMed=2988123;			
RA	Suedhof T.C., Goldstein J.L., Brown M.S., Russell D.W.;			
RT	"The LDL receptor gene: a mosaic of exons shared with different			
RT	proteins.";			
RL	Science 228:815-822(1985).			
RN	[3]			
RP	MUTAGENESIS OF CYTOPLASMIC DOMAIN.			
RX	MEDLINE=87166014; PubMed=3104336;			
RA	Davis C.G., van Driel I.R., Russell D.W., Brown M.S., Goldstein J.L.;			
RT	"The low density lipoprotein receptor. Identification of amino acids			
RT	in cytoplasmic domain required for rapid endocytosis.";			
RL	J. Biol. Chem. 262:4075-4082(1987).			
RN	[4]			
RP	O-GLYCOSYLATION DOMAIN.			
RX	MEDLINE=86140036; PubMed=3005267;			
RA	Davis C.G., Elhammer A., Russell D.W., Schneider W.J., Kornfeld S.,			
RA	Brown M.S., Goldstein J.L.;			
RT	"Deletion of clustered O-linked carbohydrates does not impair			
RT	function of low density lipoprotein receptor in transfected			
RT	fibroblasts.";			
RL	J. Biol. Chem. 261:2828-2838(1986).			
RN	[5]			
RP	STRUCTURE BY NMR OF 20-67.			
RX	MEDLINE=95327641; PubMed=7603991;			
RA	Daly N.L., Scanlon M.J., Djordjevic J.T., Kroon P.A., Smith R.;			
RT	"Three-dimensional structure of a cysteine-rich repeat from the low-			
RT	density lipoprotein receptor.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 92:6334-6338(1995).			
RN	[6]			
RP	STRUCTURE BY NMR OF 65-104.			
RX	MEDLINE=96062511; PubMed=7578052;			
RA	Daly N.L., Djordjevic J.T., Kroon P.A., Smith R.;			
RT	"Three-dimensional structure of the second cysteine-rich repeat from			

RT the human low-density lipoprotein receptor.";  
 RL Biochemistry 34:14474-14481(1995).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 196-232.  
 RX MEDLINE=97404319; PubMed=9262405;  
 RA Fass D., Blacklow S.C., Kim P.S., Berger J.M.;  
 RT "Molecular basis of familial hypercholesterolemia from structure of  
 RL LDL receptor module.";  
 RL Nature 388:691-693(1997).  
 RN [8]  
 RP REVIEW ON FH VARIANTS.  
 RX MEDLINE=93250847; PubMed=1301956;  
 RA Hobbs H.H., Brown M.S., Goldstein J.L.;  
 RT "Molecular genetics of the LDL receptor gene in familial  
 RL hypercholesterolemia.";  
 RL Hum. Mutat. 1:445-466(1992).  
 RN [9]  
 RP REVIEW ON FH VARIANTS.  
 RX MEDLINE=97169388; PubMed=9016531;  
 RA Varret M., Rabes J.-P., Colod-Beroud G., Junien J., Boilleau C.,  
 RT "Software and database for the analysis of mutations in the human LDL  
 RL receptor gene.";  
 RL Nucleic Acids Res. 25:172-180(1997).  
 RN [10]  
 RP VARIANTS FH AFRIKANER-1; -2 AND -3.  
 RX MEDLINE=89340934; PubMed=2569482;  
 RA Leitersdorf E., van der Westhuizen D.R., Coetzee G.A., Hobbs H.H.;  
 RT "Two common low density lipoprotein receptor gene mutations cause  
 RL familial hypercholesterolemia in Afrikaners.";  
 RL J. Clin. Invest. 84:954-961(1989).  
 RN [11]  
 RP VARIANT FH BARI.  
 RX MEDLINE=86161691; PubMed=3955657;  
 RA Davis C.G., Lehman M.A., Russell D.W., Anderson R.G.W., Brown M.S.,  
 RL Goldstein J.L.;  
 RT "The J.D. mutation in familial hypercholesterolemia: amino acid  
 RL substitution in cytoplasmic domain impedes internalization of LDL  
 RL receptors.";  
 RL Cell 45:15-24(1986).  
 RN [12]  
 RP VARIANTS FH DURBAN-1 AND -2.  
 RX MEDLINE=93350003; PubMed=8347689;  
 RA Rubinstein D.C., Jialal I., Leitersdorf E., Coetzee G.A.,  
 RL van der Westhuizen D.R.;  
 RT "Identification of two new LDL-receptor mutations causing homozygous  
 RL familial hypercholesterolemia in a South African of Indian origin.";  
 RL Biochim. Biophys. Acta 1182:75-82(1993).  
 RN [13]  
 RP VARIANTS FH FRENCH CANADIAN-2; -3 AND -4.  
 RX MEDLINE=90203205; PubMed=2318961;  
 RA Leitersdorf E., Tobin E.J., Davignon J., Hobbs H.H.;  
 RT "Common low-density lipoprotein receptor mutations in the French  
 RL Canadian population.";  
 RL J. Clin. Invest. 85:1014-1023(1990).  
 RN [14]  
 RP VARIANT FH OSAKA-3.  
 RX MEDLINE=93076772; PubMed=1446662;  
 RA Miyake Y., Tajima S., Funahashi T., Yamamura T., Yamamoto A.;  
 RT "A point mutation of the low-density lipoprotein receptor causing rapid  
 RL degradation of the receptor.";  
 RL Eur. J. Biochem. 210:1-7(1992).  
 RN [15]  
 RP VARIANT FH PISCATAMAY.  
 RX MEDLINE=91328152; PubMed=1867200;  
 RA Meiner V., Landsberger D., Berkman N., Reshef A., Segal P.,  
 RL Seftel H.C., van der Westhuizen D.R., Jeenan M.S., Coetzee G.A.,  
 RT "A common lithuanian mutation causing familial hypercholesterolemia  
 RL in Ashkenazi Jews.";  
 RL Am. J. Hum. Genet. 49:443-449(1991).  
 RN [16]  
 RP VARIANT FH SAFED.

RX MEDLINE=93216279; PubMed=8462973;  
 RA Leitersdorf E., Reshef A., Meiner V., Dann E.J., Beigel Y.,  
 RA van Rogen F.G., van der Westhuizen D.R., Coetzee G.A.;  
 RT "A missense mutation in the low density lipoprotein receptor gene  
 RL causes familial hypercholesterolemia in Sephardic Jews.";  
 RL Hum. Genet. 91:141-147(1993).  
 RN [17]  
 RP VARIANT FH TRIESTE.  
 RX MEDLINE=94222419; PubMed=8166830;  
 RA Tellì N., Garuti R., Pedrazzi P.,  
 RA Tiozzo R., Catlin L., Valenti M., Rolletti M., Bertolini S.,  
 RA Stefanutti C., Calandra S.;  
 RT "A new missense mutation (Cys297-->Phe) of the low density  
 RL lipoprotein receptor in Italian patients with familial  
 RL hypercholesterolemia (FHTrieste).";  
 RL Hum. Genet. 93:538-540(1994).  
 RN [18]  
 RP VARIANT FH ZAMBIA.  
 RX MEDLINE=89264579; PubMed=2726768;  
 RA Soutar A.K., Knight B.L., Patel D.D.;  
 RT "Identification of a point mutation in growth factor repeat C of the  
 RL low density lipoprotein-receptor gene in a patient with homozygous  
 RL familial hypercholesterolemia that affects ligand binding and  
 RL intracellular movement of receptors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4166-4170(1989).  
 RN [19]  
 RP VARIANT FH ZAMBIA.  
 RX MEDLINE=93100522; PubMed=1464748;  
 RA Rubinstein D.C., Coetzee G.A., Marais A.D., Leitersdorf E.,  
 RA Seftel H.C., van der Westhuizen D.R.;  
 RT "Identification and properties of the proline664-leucine mutant LDL  
 RL receptor in South Africans of Indian origin.";  
 RL J. Lipid Res. 33:1647-1653(1992).  
 RN [20]  
 RP VARIANTS FH PORI HIS-401 AND TURKU ASP-844.  
 RX MEDLINE=96029270; PubMed=7573037;  
 RA Koivisto U.-M., Viikari J.S., Kontula K.;  
 RT "Molecular characterization of minor gene rearrangements in Finnish  
 RL patients with heterozygous familial hypercholesterolemia:  
 RL identification of two common missense mutations (Gly823-->Asp and  
 RL Leu380-->His) and eight rare mutations of the LDL receptor gene.";  
 RL Am. J. Hum. Genet. 57:789-797(1995).  
 RN [21]  
 RP VARIANTS FH LYS-140; SER-338 AND LEU-685.  
 RX MEDLINE=96011600; PubMed=7583548;  
 RA Maruyama T., Miyake Y., Tajima S., Harada-Shiba M., Yamamura T.,  
 RA Tsuchida M., Kishino B.-I., Horiguchi Y., Funahashi T., Matsuzawa Y.,  
 RA Yamamoto A.;  
 RT "Common mutations in the low-density lipoprotein-receptor gene  
 RL causing familial hypercholesterolemia in the Japanese population.";  
 RL Arterioscler. Thromb. Vasc. Biol. 15:1713-1718(1995).  
 RN [22]  
 RP VARIANT FH FRENCH HIS-564.  
 RX MEDLINE=96055524; PubMed=7550239;  
 RA Triot-Guerber F., Saint-Jore B., Valenti K., Foulon T., Bost M.,  
 RA Hadjian A.J.;  
 RT "Identification of a mutation, N543H, in exon 11 of the low-density  
 RL lipoprotein receptor gene in a French family with familial  
 RL hypercholesterolemia.";  
 RL Hum. Mutat. 6:87-88(1995).  
 RN [23]  
 RP VARIANTS FH SWEDISH LYS-277; THR-423 AND ASN-579.  
 RX MEDLINE=95362239; PubMed=7635461;  
 RA Ekstrom U., Adrhamson M., Sveger T., Lombardi P., Nilsson-Ehle P.;  
 RT "An efficient screening procedure detecting six novel mutations in  
 RL the LDL receptor gene in Swedish children with  
 RL hypercholesterolemia.";  
 RL Hum. Genet. 96:147-150(1995).  
 RN [24]  
 RP VARIANT FH NORWEGIAN ASN-487 DEL.  
 RX MEDLINE=95362260; PubMed=7635482;  
 RA Ieren T.P., Solberg K., Rodningsen O.K., Tonstad S., Ose L.;  
 RT "Two novel point mutations in the EGF precursor homology domain of

Query Match 78.6%; Score 33; DB 1; Length 860;  
 Best Local Similarity 87.5%; Pred. No. 19;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLAEE 8  
 DB 637 SDYNLAEE 644

## RESULT 2

NIFE\_RHIME STANDARD: PRT: 476 AA.  
 AC 092AT2;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE Nitrogenase iron-molybdenum cofactor biosynthesis protein nife.  
 GN NIFE OR RA0452 OR SMA0830.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSyma (megaplasmid 1).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=1021;  
 RX MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouy J.,  
 RA Gurjal M., Hong A., Hutzar L., Hyman R.W., Kahn M.L.,  
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSyma megaplasmid."  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 CC -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF  
 CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).  
 CC -1- PATHWAY: FE-MO COFACTOR BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AE007236; AAK65110.1;  
 DR PROSITE: PS00699; NITROGENASE\_1.1;  
 DR PROSITE: PS00090; NITROGENASE\_1.2;  
 KW Nitrogen fixation; Plasmid; Complete proteome.  
 SQ SEQUENCE 476 AA; 52358 MW; 83CECE94B4B1AC713 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 476;  
 Best Local Similarity 55.6%; Pred. No. 17;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEE 9  
 DB 195 SDYNLAEE 203

RESULT 3  
 HEX\_ADEG1 STANDARD: PRT: 942 AA.

AC P42671; 064758;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hexon protein (late protein 2).  
 DB PII.

OS Avian adenovirus gall (strain Pheips) (Fowl adenovirus 1) (CELO).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.  
 OX NCBI\_TaxID=10553;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=96186720; PubMed=8627769;  
 RA Chlocca S., Kurzbauer R., Schaffner G., Baker A., Mautner V.,  
 RA Cotten M.;  
 RT "The complete DNA sequence and genomic organization of the avian  
 RT adenovirus CELO."  
 RL J. Virol. 70:2939-2949(1996).  
 RN [2]

RP SEQUENCE OF 913-942 FROM N.A.  
 RX MEDLINE=93362429; PubMed=8395124;  
 RA Cai F., Weber J.M.;  
 RT "Organization of the avian adenovirus genome and the structure of its  
 RT endoplasmic reticulum."  
 RT Virology 196:358-362(1993).  
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE STRUCTURAL PROTEINS IN THE  
 CC VIRAL COAT AND IS SYNTHESIZED DURING LATE INFECTION (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: U46933; AAC54912.1;  
 DR EMBL: L13161; AAA51401.1;  
 DR HSSP: P03327; IDHX  
 DR InterPro: IPR000736; Adeno\_hexon.  
 DR Pfam: PF01065; Adeno\_hexon; 1.  
 DR ProDom: PD002815; Adeno\_hexon; 1.  
 KW Coat protein; Hexon protein; Late protein.  
 SQ SEQUENCE 942 AA; 106709 MW; 7F4CE8D5F17D051B CRC64;

Query Match 76.2%; Score 32; DB 1; Length 942;  
 Best Local Similarity 66.7%; Pred. No. 36;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEE 9  
 DB 581 TEVNLMANF 589

## RESULT 4

ORN\_VIBCH STANDARD: PRT: 181 AA.  
 AC 09KV17;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Oligoribonuclease (EC 3.1.-.-).  
 GN ORN OR VC0341.

OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibri.  
 OX NCBI\_TaxID=666;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickley E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dracot I., Sellers P.,  
 RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

```

RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: 3'-TO-5' EXORIBONUCLEASE SPECIFIC FOR SMALL
CC OLIGORIBONUCLEOTIDES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE OLIGORIBONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004123; AAF93514.1; -.
DR TIGR: VC0341; -.
DR InterPro: IPR000520; Exonuclease.
DR Pfam: PF00929; Exonuclease; 1.
DR SMART: SM00479; EXOIII; 1.
DR Hydrolase: Exonuclease; Nuclease; Complete proteome.
KW ACT_SITE 129 129 POTENTIAL.
FT ACCT_SITE 129 129 AAC071B9C94FEFE3 CRC64;
SQ SEQUENCE 181 AA; 20929 MW; AAC071B9C94FEFE3 CRC64;

Query Match 73.8%; Score 31; DB 1; Length 181;
Best Local Similarity 75.0%; Pred. No. 9.9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLIAE 8
DB 35 SEVNLIAE 42

RESULT 5
NEK2_MOUSE STANDARD; PRT; 443 AA.
AC 035942: 035959;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Serine/threonine-protein kinase NEK2 (EC 2.7.1.1.-) (NIMA-related
DE protein kinase 2).
GN NEK2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RC STRAIN=SWISS WEBSTER; TISSUE=Testis;
RX MEDLINE=97330684; PubMed=9187143;
RA Rhee K., Wolgemuth D.J.;
RT "The NIMA-related kinase 2, Nek2, is expressed in specific stages of
RT the meiotic cell cycle and associates with meiotic chromosomes.";
RL Development 124:2167-2177(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98243037; PubMed=9583679;
RA Arama E., Janai A., Killin G., Motro B.;
RT "Murine NIMA-related kinases are expressed in patterns suggesting
RT distinct functions in gametogenesis and a role in the nervous
RT system.";
RL Oncogene 16:1813-1823(1998).
RN [3]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Brain;
RX MEDLINE=98096235; PubMed=9434622;
RA Tanaka K., Parvinen M., Nigg E.A.;
RT "The in vivo expression pattern of mouse Nek2, a NIMA-related kinase,
RT indicates a role in both mitosis and meiosis.";
RL Exp. Cell Res. 237:264-274(1997).
CC -1- FUNCTION: PROTEIN KINASE THAT IS INVOLVED IN MITOTIC REGULATION.
CC MAY HAVE A ROLE AT THE G2-M TRANSITION. MAY ALSO PLAY A ROLE IN

```

```

CC MEIOSIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANTLY EXPRESSED IN TESTIS. LOW
CC LEVELS FOUND IN MID-GESTATION EMBRYO, OVARY, PLACENTA, INTESTINE,
CC THYMUS AND SKIN. WITHIN THE TESTIS, EXPRESSION RESTRICTED TO GERM
CC CELLS WITH HIGHEST LEVELS DETECTED IN SPERMATOCYTES AT PACHYENE
CC AND DIPTERENE STAGES. ALSO EXPRESSED IN MEIOTIC PACHYENE OOCYTES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC NIMA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U95610; AAB67973.1; -.
DR EMBL: AF013166; AAC35393.1; -.
DR EMBL: AF007247; AAB70470.1; -.
DR HSSP: P24941; ICKP.
DR MGD: MGI:109359; Nek2.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW Nuclear protein; Phosphorylation; Cell cycle; Cell division; Meiosis.
FT DOMAIN 8 271 ATP (BY SIMILARITY).
FT NR_BIND 14 22 ATP (BY SIMILARITY).
FT BINDING 37 37 ATP (BY SIMILARITY)..
FT ACT_SITE 141 141 BY SIMILARITY.
FT MOD_RES 179 179 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 443 AA; 51307 MW; DE09565C378307E1 CRC64;

Query Match 73.8%; Score 31; DB 1; Length 443;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLIAE 8
DB 54 SEVNLIAE 61

RESULT 6
NEK2_HUMAN STANDARD; PRT; 445 AA.
AC P51955;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 01-OCT-1996 (Rel. 34; Last annotation update)
DE 16-OCT-2001 (Rel. 40; Last sequence update)
DE Serine/threonine-protein kinase NEK2 (EC 2.7.1.1.-) (NIMA-related
DE protein kinase 2) (NIMA-like protein kinase 1) (HSK 21).
GN NEK2 OR NLR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cell, Placenta, and Nasopharynx;
RX MEDLINE=94368699; PubMed=7522034;
RA Schultz S.J., Fry A.M., Suetterlin C., Ried T., Nigg E.A.;
RT "Cell cycle-dependent expression of Nek2, a novel human protein
RT kinase related to the NIMA mitotic regulator of Aspergillus
RT nidulans.";
RL Cell Growth Differ. 5:625-635(1994).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA Lu K.P., Hunter T.;
RT "Molecular cloning and expression of NUK1, a human NIMA-like kinase.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 83-203 FROM N.A.
RX MEDLINE-94100173; PubMed-8274451;
RA Schultz S.J., Migg E.A.;
RT "Identification of 21 novel human protein kinases, including 3 members
of a family related to the cell cycle regulator nima of Aspergillus
nidulans.";
RL Cell Growth Differ. 4:821-830(1993).
CC -1- FUNCTION: PROTEIN KINASE THAT IS INVOLVED IN MITOTIC REGULATION.
CC MAY HAVE A ROLE AT THE G2-M TRANSITION. MAY ALSO PLAY A ROLE IN
CC MEIOSIS.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES THROUGHOUT S PHASE AND SHOWS
CC MAXIMAL LEVELS IN LATE G2. THIS EXPRESSION PATTERN IS HIGHLY
CC REMINISCENT OF THAT OF A AND B CYCLINS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC NIMA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29066; CA82309.1; -;
DR EMBL; U11050; AAI19558.1; -;
DR EMBL; Z25425; CA80912.1; -;
DR HSSP; P00518; 1PHK.
DR MIM; 604043; -;
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002290; Ser_Thr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Mitosis;
KW Nuclear protein; Phosphorylation; Cell cycle; Cell division; Meiosis.
FT DOMAIN 8 271 PROTEIN KINASE.
FT NP_BIND 14 22 ATP (BY SIMILARITY).
FT BINDING 37 37 ATP (BY SIMILARITY).
FT ACT_SITE 141 141 BY SIMILARITY.
FT MOD_RES 179 179 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 84 85 IV -> LY (IN REF. 3).
SQ SEQUENCE 445 AA; 51763 MW; D33A3778ABBD9E CRC64;

Query Match 73.8%; Score 31; DB 1; Length 445;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLAE 8
DB 54 SEVNLAE 61

RESULT 7
Y041_AQUAE STANDARD; PRT; 103 AA.
ID Y041_AQUAE
AC 066457;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_041.
GN AQ_041.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID-63363;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE-98196666; PubMed-9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.";
RL Nature 392:353-358(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0102 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000671; AAC06429.1; -;
DR InterPro; IPR003509; UPF0102.
DR Pfam; PF02021; UPF0102; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11870 MW; 78E285829D39F87E CRC64;

Query Match 71.4%; Score 30; DB 1; Length 103;
Best Local Similarity 62.5%; Pred. No. 9;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 EVNLAEF 9
DB 32 EIDLAEF 39

RESULT 8
Y778_AQUAE STANDARD; PRT; 328 AA.
ID Y778_AQUAE
AC 066974;
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein AQ_778.
GN AQ_778.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID-63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE-98196666; PubMed-9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.";
RL Nature 392:353-358(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0052 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000707; AAC06942.1; -;
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR002882; UPF0052.
DR Pfam; PF01933; UPF0052; 1.
DR PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 328 AA; 35957 MW; 30BB21AE995958F8B CRC64;

```

```

Query Match      71.4%; Score 30; DB 1; Length 328;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      3 VNLAEEF 9
      |||||
DB      133 VNLAEEF 139

RESULT 9
ID      SCHPO
ID      VD15_SCHPO      STANDARD;      PRT;      581 AA.
AC      Q10238;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Putative regulatory protein CAG9.05.
GN      SPAC49.05.
OS      Schizosaccharomyces pombe (fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC      Schizosaccharomycetales; Schizosaccharomycetaceae;
OC      Schizosaccharomycetes.
OX      NCBI_TaxID=4896;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=972;
RA      Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL      Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: SOME, TO DROSOPHILA PUM.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; 269727; CAA93555.1; -.
DR      InterPro; IPR001313; PDM.
DR      Pfam; PF00806; PUF; 6.
DR      SMART; SM00025; Pumilio; 7.
DR      SMART; SM00025; Pumilio; 7.
KW      Hypothetical protein.
SQ      SEQUENCE      581 AA; 66866 MW; 93A360BA7BD0A40E CRC64;

Query Match      71.4%; Score 30; DB 1; Length 581;
Best Local Similarity 62.5%; Pred. No. 59;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY      1 SEVNLAEE 8
      |||||
DB      547 SQINLAEE 554

RESULT 10
MECA_CAEEL
ID      MECA_CAEEL      STANDARD;      PRT;      768 AA.
AC      P24612;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Mechanosensory protein 4 (Degenerin mec-4).
GN      MEC-4 OR MEC-13 OR TOLC8.7.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN-BRISTOL N2;
RX      MEDLINE=96251674; PubMed=8655580;
RA      Lai C.C., Hong K., Kinnell M., Driscoll M.;

```

```

RT      "Sequence and transmembrane topology of MEC-4, an ion channel subunit
RT      required for mechanotransduction in Caenorhabditis elegans.";
RL      J Cell Biol. 133:1071-1081(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BRISTOL N2;
RA      Wohlmann P., Hawkins J.;
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 271-768 FROM N.A.
RC      STRAIN-BRISTOL N2;
RX      MEDLINE=91156026; PubMed=1672038;
RA      Driscoll M., Chalfie M.;
RT      "The mec-4 gene is a member of a family of Caenorhabditis elegans
RT      genes that can mutate to induce neuronal degeneration.";
RL      Nature 349:588-593(1991).
CC      -1- FUNCTION: PROBABLE SODIUM CHANNEL SUBUNIT. MAY BE NEEDED FOR
CC      MECHANOSENSORY TRANSDUCTION (TOUCH SENSITIVITY).
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -1- DISEASE: MUTATIONS IN MEC-4 RESULTS IN THE DEGENERATION OF A SMALL
CC      SET OF NEURONS WHICH TYPICALLY SWEET TO SEVERAL TIME THEIR NORMAL
CC      DIAMETER BEFORE THEY DISAPPEAR. PRESUMABLY DUE TO LYSIS.
CC      -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U53669; AAC47265.1; -.
DR      EMBL; U58726; AAB00580.1; ALT_INIT.
DR      EMBL; X58982; CAA41731.1; -.
DR      PIR; S13645; S13645.
DR      WormPep; T01C8.7; CE07463.
DR      InterPro; IPR001873; ASC.
DR      Pfam; PF00858; ASC; 1.
DR      PRINTS; PR01078; AMINACHANNEL.
DR      PROSITE; PS01206; ASC; 1.
KW      Ionic channel; Transmembrane; Ion transport; Glycoprotein;
KW      Neutrogenesis.
SQ      SEQUENCE      768 AA; 87194 MW; 8899D0FCB7F02C6 CRC64;

Query Match      71.4%; Score 30; DB 1; Length 768;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      3 VNLAEEF 9
      |||||
DB      709 VNLAEEF 715

RESULT 11
MECA_CAEER
ID      MECA_CAEER      STANDARD;      PRT;      769 AA.

```



AC 017298;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Mechanosensory protein 4 (Degenerin mec-4).  
 GN MEC-4.  
 OS Caenorhabditis briggsae.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6238;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96251674; PubMed=8655580;  
 RA Lai C.C., Hong K., Kinnell M., Chalfie M., Driscoll M.;  
 RT "Sequence and transmembrane topology of MEC-4, an ion channel subunit  
 required for mechanotransduction in *Caenorhabditis elegans*.";  
 RL J. Cell Biol. 133:1071-1081(1996).  
 CC -1- FUNCTION: PROBABLE SODIUM CHANNEL SUBUNIT. MAY BE NEEDED FOR  
 CC MECHANOSENSORY TRANSDUCTION (TOUCH SENSITIVITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS  
 CC FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: U53670; AAC47264.1; -  
 DR InterPro: IPR001873; ASC.  
 DR Pfam: PF00858; ASC. 1.  
 DR PRINTS: PR01078; AMINACHANNEL.  
 DR PROSITE: PS01206; ASC. 1.  
 KW Ionic channel; Transmembrane; Ion transport; Glycoprotein;  
 KM Neurodegeneration.  
 FT DOMAIN 1 109  
 FT TRANSMEM 110 130 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 131 719 POTENTIAL.  
 FT TRANSMEM 720 740 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 741 769 POTENTIAL.  
 FT CARBOHYD 336 336 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 769 AA; 87141 MW; 430BB718BA0F5002 CRC64;  
 Query Match 71.4%; Score 30; DB 1; Length 769;  
 Best Local Similarity 85.7%; Pred. NO. 80;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 VNLAEF 9  
 DB 710 VNLADF 716  
 RESULT 12  
 LDLR\_CRIGR STANDARD: PRT: 854 AA.  
 AC P35950;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Low-density lipoprotein receptor precursor (LDL receptor).  
 GN LDLR OR LDLA.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.

OX NCBI\_TaxID=10029;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92407468; PubMed=1527478;  
 RA Bishop R.W.;  
 RT "Structure of the hamster low density lipoprotein receptor gene.";  
 RL J. Lipid Res. 33:549-557(1992).  
 RN (2)  
 RP SEQUENCE OF 570-615 FROM N.A.  
 RX MEDLINE=87064645; PubMed=3785227;  
 RA Sege R.D., Kozarsky K.F., Krieger M.;  
 RT "Characterization of a family of gamma-ray-induced CHO mutants  
 demonstrates that the LDLA locus is diploid and encodes the  
 low-density lipoprotein receptor.";  
 RL Mol. Cell. Biol. 6:3268-3277(1986).  
 CC -1- FUNCTION: BINDS LDL, THE MAJOR CHOLESTEROL-CARRYING LIPOPROTEIN  
 CC OF PLASMA, AND TRANSPORTS IT INTO CELLS BY ENDOCYTOSIS. IN ORDER  
 CC TO BE INTERNALIZED, THE RECEPTOR-LIGAND COMPLEXES MUST FIRST  
 CC CLUSTER INTO CLATHRIN-COATED PITS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 6 LDL-RECEPTOR CLASS B DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: M94387; AA51449.1; -  
 DR EMBL: M18877; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A48908; ORHYLD.  
 DR HSSP: P01130; ILDR.  
 DR InterPro: IPR000152; Asx hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-like.  
 DR InterPro: IPR002172; LDL\_recept\_A.  
 DR InterPro: IPR000033; LDL\_receptor\_rep.  
 DR Pfam: PF00008; EGF. 3.  
 DR Pfam: PF00057; Idl\_recept\_a; 7.  
 DR Pfam: PF00058; Idl\_recept\_b; 5.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR SMART: SM00179; EGF\_CA\_1.  
 DR SMART: SM00001; EGF\_Like; 2.  
 DR SMART: SM00192; LDLA; 7.  
 DR SMART: SM00135; LY; 5.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01165; EGF\_2; 2.  
 DR PROSITE: PS01187; EGF\_CA\_1.  
 DR PROSITE: PS01209; LDLRA\_1; 7.  
 DR PROSITE: PS50068; LDLRA\_2; 7.  
 KW Glycoprotein; LDL; Cholesterol metabolism; Lipid transport;  
 KW Endocytosis; Coated pits; Transmembrane; Receptor; Signal;  
 KW EGF-like domain; Repeat.  
 FT SIGNAL 1 21  
 FT CHAIN 22 854  
 FT DOMAIN 22 782  
 FT TRANSMEM 783 804  
 FT DOMAIN 805 854  
 FT DOMAIN 25 316  
 FT DOMAIN 25 65  
 FT DOMAIN 66 106  
 FT DOMAIN 107 145  
 FT DOMAIN 146 186  
 FT DOMAIN 196 234  
 FT DOMAIN 235 273  
 FT DOMAIN 275 314  
 FT DOMAIN 315 354  
 FT EGF-LIKE 1.  
 -----  
 FT POTENTIAL.  
 FT LOW-DENSITY LIPOPROTEIN RECEPTOR.  
 FT EXTRACELLULAR (POTENTIAL).  
 FT POTENTIAL.  
 FT CYTOPLASMIC (POTENTIAL).  
 FT 7 X 41 AA TANDEN REPEATS OF LDL-RECEPTOR  
 FT CLASS DOMAIN; LIGAND BINDING.  
 FT LDL-RECEPTOR CLASS A 1.  
 FT LDL-RECEPTOR CLASS A 2.  
 FT LDL-RECEPTOR CLASS A 3.  
 FT LDL-RECEPTOR CLASS A 4.  
 FT LDL-RECEPTOR CLASS A 5.  
 FT LDL-RECEPTOR CLASS A 6.  
 FT LDL-RECEPTOR CLASS A 7.  
 FT EGF-LIKE 1.

```

FT DOMAIN 355 394 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 398 439 LDL-RECEPTOR CLASS B 1.
FT REPEAT 440 485 LDL-RECEPTOR CLASS B 2.
FT REPEAT 486 528 LDL-RECEPTOR CLASS B 3.
FT REPEAT 529 572 LDL-RECEPTOR CLASS B 4.
FT REPEAT 573 615 LDL-RECEPTOR CLASS B 5.
FT REPEAT 616 657 LDL-RECEPTOR CLASS B 6.
FT DOMAIN 663 712 EGF-LIKE 3.
FT DOMAIN 721 765 CLUSTERED O-LINKED OLIGOSACCHARIDES.
FT SITE 817 822 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT DISULFID 27 39 BY SIMILARITY.
FT DISULFID 34 52 BY SIMILARITY.
FT DISULFID 46 63 BY SIMILARITY.
FT DISULFID 68 82 BY SIMILARITY.
FT DISULFID 75 95 BY SIMILARITY.
FT DISULFID 89 104 BY SIMILARITY.
FT DISULFID 109 121 BY SIMILARITY.
FT DISULFID 116 134 BY SIMILARITY.
FT DISULFID 128 143 BY SIMILARITY.
FT DISULFID 148 160 BY SIMILARITY.
FT DISULFID 155 173 BY SIMILARITY.
FT DISULFID 167 184 BY SIMILARITY.
FT DISULFID 198 210 BY SIMILARITY.
FT DISULFID 205 223 BY SIMILARITY.
FT DISULFID 217 232 BY SIMILARITY.
FT DISULFID 237 249 BY SIMILARITY.
FT DISULFID 244 262 BY SIMILARITY.
FT DISULFID 256 271 BY SIMILARITY.
FT DISULFID 277 290 BY SIMILARITY.
FT DISULFID 285 303 BY SIMILARITY.
FT DISULFID 297 314 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 326 339 BY SIMILARITY.
FT DISULFID 341 353 BY SIMILARITY.
FT DISULFID 359 369 BY SIMILARITY.
FT DISULFID 365 378 BY SIMILARITY.
FT DISULFID 380 393 BY SIMILARITY.
FT DISULFID 667 681 BY SIMILARITY.
FT DISULFID 677 696 BY SIMILARITY.
FT DISULFID 698 711 BY SIMILARITY.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 854 AA; 94516 MW; 5D50EBFBD834B35E CRC64;

Query Match 71.4%; Score 30; DB 1; Length 854;
Best Local Similarity 75.0%; Pred. No. 90;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 SEVNLAE 8
DB 637 SDVNLVAE 644

RESULT 13
LDLR_MOUSE STANDARD; PRT; 864 AA.
AC P35951;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Low-density lipoprotein receptor precursor (LDL receptor).
GN LDLR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=BAIB/C; TISSUE=Liver;
RX MEDLINE=93221531; Pubmed=8466528;
RA Hofter M.J.V., van Eck M.M., Petitj F., van der Zee A., de Wit E.,
RA Meijer D., Grosveld G., Havekes L.M., Hofker M.H., Frans R.R.;

```

```

RT "The mouse low density lipoprotein receptor gene: cDNA sequence and
RT exon-intron structure."
RT Biochem. Biophys. Res. Commun. 191:880-886(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93117934; Pubmed=1475710;
RA Polyano W.J., Dichek D.A., Mason J., Anderson W.F.;
RT "Molecular cloning and nucleotide sequence of cDNA encoding a
RT functional murine low density lipoprotein receptor."
RT Somat. Cell Mol. Genet. 18:443-450(1992).
RL -1- FUNCTION: BINDS LDL, THE MAJOR CHOLESTEROL-CARRYING LIPOPROTEIN
CC OF PLASMA, AND TRANSPORTS IT INTO CELLS BY ENDOCYTOSIS. IN ORDER
CC TO BE INTERNALIZED, THE RECEPTOR-LIGAND COMPLEXES MUST FIRST
CC CLUSTER INTO CLATHRIN-COATED PITS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 LDL-RECEPTOR CLASS B DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z19521; CAA79581.1; -
DR EMBL: X64414; CAA45759.1; -
DR PIR: JN0461; ORMSLD.
DR HSSP: P01130; LDLR.
DR MGD: MGI:96765; LDLR.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-Like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR000033; LDL_recept_rep.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00057; ldl_recept_a; 7.
DR Pfam: PF00058; ldl_recept_b; 5.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_Like; 2.
DR SMART: SM00192; LDLA; 7.
DR SMART: SM00135; LY; 5.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01209; LDLRA_1; 7.
DR PROSITE: PS50068; LDLRA_2; 7.
DR GlycoProtex: LDL; Cholesterol metabolism; Lipid transport;
KW Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
KW EGF-like domain; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 864 LOW-DENSITY LIPOPROTEIN RECEPTOR.
FT DOMAIN 22 790 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 791 812 POTENTIAL.
FT DOMAIN 813 864 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 25 65 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 66 106 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 107 145 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 146 186 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 196 234 LDL-RECEPTOR CLASS A 5.
FT DOMAIN 235 273 LDL-RECEPTOR CLASS A 6.
FT DOMAIN 275 314 LDL-RECEPTOR CLASS A 7.
FT DOMAIN 315 354 EGF-LIKE 1.
FT DOMAIN 355 394 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 398 439 LDL-RECEPTOR CLASS B 1.
FT REPEAT 440 485 LDL-RECEPTOR CLASS B 2.
FT REPEAT 486 528 LDL-RECEPTOR CLASS B 3.
FT REPEAT 529 572 LDL-RECEPTOR CLASS B 4.

```

```

FT REPEAT 573 615 LDL-RECEPTOR CLASS B 5.
FT REPEAT 616 657 LDL-RECEPTOR CLASS B 6.
FT DOMAIN 663 713 EGF-LIKE 3.
FT SITE 722 770 CLUSTERED O-LINKED OLIGOSACCHARIDES.
FT SITE 827 832 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT DISULFID 34 52 BY SIMILARITY.
FT DISULFID 46 63 BY SIMILARITY.
FT DISULFID 68 82 BY SIMILARITY.
FT DISULFID 75 95 BY SIMILARITY.
FT DISULFID 89 104 BY SIMILARITY.
FT DISULFID 109 121 BY SIMILARITY.
FT DISULFID 116 134 BY SIMILARITY.
FT DISULFID 128 143 BY SIMILARITY.
FT DISULFID 148 160 BY SIMILARITY.
FT DISULFID 155 173 BY SIMILARITY.
FT DISULFID 167 184 BY SIMILARITY.
FT DISULFID 198 210 BY SIMILARITY.
FT DISULFID 205 223 BY SIMILARITY.
FT DISULFID 217 232 BY SIMILARITY.
FT DISULFID 237 249 BY SIMILARITY.
FT DISULFID 244 262 BY SIMILARITY.
FT DISULFID 256 271 BY SIMILARITY.
FT DISULFID 277 290 BY SIMILARITY.
FT DISULFID 285 303 BY SIMILARITY.
FT DISULFID 297 314 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 326 339 BY SIMILARITY.
FT DISULFID 341 353 BY SIMILARITY.
FT DISULFID 359 369 BY SIMILARITY.
FT DISULFID 365 378 BY SIMILARITY.
FT DISULFID 380 393 BY SIMILARITY.
FT DISULFID 667 682 BY SIMILARITY.
FT DISULFID 678 697 BY SIMILARITY.
FT DISULFID 699 712 BY SIMILARITY.
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 23 23 V -> A (IN REF. 2).
FT CONFLICT 27 27 G -> C (IN REF. 2).
FT CONFLICT 61 61 K -> E (IN REF. 2).
FT CONFLICT 144 144 P -> Q (IN REF. 2).
FT CONFLICT 156 156 K -> N (IN REF. 2).
FT CONFLICT 178 178 H -> D (IN REF. 2).
FT CONFLICT 186 187 AE -> GR (IN REF. 2).
FT CONFLICT 820 821 MISSING (IN REF. 2).
SQ SEQUENCE 864 AA: 95134 MW: 4AE15953ED11E0E CRC64;

Query Match
Best Local Similarity 71.4%; Score 30; DB 1; Length 864;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SEVNLAE 8
Db 637 SDYNLVAE 644

RESULT 14
ID IF2_CHLPN STANDARD; PRT: 890 AA.
AC Q928M1; Q9JRX1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Translation initiation factor IF-2.
GN INF8 OR CPN0317 OR CP0440.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
Kalmán S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

```

```

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson M., Deyou R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shital M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
CC OF PROTEIN SYNTHESIS IN VITRO, PROTECTS FORMYL METHIONYL-TRNA FROM
CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
CC RIBOSOMAL SUBUNIT. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001616; AAD18466.1; -
DR EMBL: AE002205; AAF38279.1; -
DR EMBL: AP002546; BAA98527.1; -
DR HSSP: P02990; IFTU.
DR TIGR: CP0440; -
DR InterPro: IPR000795; GTP_EFTU.
DR InterPro: IPR004161; GTP_EFTU_D2.
DR InterPro: IPR000178; IF2.
DR Pfam: PF00009; GTP_EFTU; 1.
DR Pfam: PF03144; GTP_EFTU_D2; 2.
DR PRINTS: PR00315; ELONGATNFCF.
DR PRODOM: PD186100; IF2; 1.
DR PROSITE: PS01176; IF2; 1.
DR KMW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 398 546 G-DOMAIN.
FT NP_BIND 404 411 GTP (BY SIMILARITY).
FT NP_BIND 450 454 GTP (BY SIMILARITY).
FT NP_BIND 504 507 GTP (BY SIMILARITY).
FT CONFLICT 321 321 T -> A (IN REF. 1).
SQ SEQUENCE 890 AA: 97104 MW: BB84C2C269268781 CRC64;

Query Match
Best Local Similarity 71.4%; Score 30; DB 1; Length 890;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLAE 8
Db 521 SEINLPE 528

RESULT 15
Y549_RICPR

```

```

ID Y549_RICPR STANDARD; PRT; 135 AA.
AC 092D01;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP549.
GN RP549.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RC MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT *The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.*;
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ235272; CAIL4998.1; -
CC DR Hypothetical protein; Complete proteome.
CC KW SEQUENCE 135 AA; 16051 MW; 210CB79EF2127C6 CRC64;
SQ
Query Match 69.0%; Score 29; DB 1; Length 135;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 EVNIIAE 8
|:|||||
Db 33 ELNIIAE 39

```

Search completed: October 30, 2002, 12:27:54  
 Job time : 2.77396 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:23:30 ; Search time 1.72482 seconds  
(without alignments)  
501.389 Million cell updates/sec

Title: US-09-724-571-91

Perfect score: 42

Sequence: 1 SEVNLLAEF 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	81.0	529	T42584	tegument protein 4
2	33	78.6	668	B96999	methy1-accepting c
3	33	78.6	860	QRHULD	LDL receptor precu
4	32	76.2	132	G72351	conserved hypotnet
5	32	76.2	279	T10882	1-aminocyclopropan
6	32	76.2	422	T09120	basic leucine z1pp
7	32	76.2	467	T10854	1-aminocyclopropan
8	32	76.2	472	T10889	1-aminocyclopropan
9	32	76.2	476	D95318	N1fe oxidoreductas
10	31	73.8	181	E82333	oligoribonuclease
11	31	73.8	225	T17795	hypothetical prote
12	31	73.8	236	AD2830	conserved hypotnet
13	31	73.8	240	D97588	hypothetical prote
14	31	73.8	394	T87588	hypothetical prote
15	31	73.8	411	T15020	probable porphyrin
16	31	73.8	435	T39719	beta transducin -
17	31	73.8	445	G01452	NTMA-like protein
18	31	73.8	461	A27672	hypothetical gag p
19	31	73.8	3169	T00286	toxin B - Escheric
20	30	71.4	103	T70303	hypothetical prote
21	30	71.4	133	E86796	arsenate reductase
22	30	71.4	234	AG1167	hypothetical prote
23	30	71.4	328	C70368	conserved hypotnet
24	30	71.4	334	AC3336	thiosulfate-bindin
25	30	71.4	342	AG1851	hypothetical prote
26	30	71.4	446	B83033	probable MFS trans
27	30	71.4	457	A95194	cell division prot
28	30	71.4	457	F98060	cell division prot
29	30	71.4	475	AF0687	bacteriophage tail

30	30	71.4	483	2	A67583	peptidoglycan bind
31	30	71.4	581	2	T38864	probable regulator
32	30	71.4	694	2	A12492	hypothetical prote
33	30	71.4	749	2	T29859	mechanosensory pro
34	30	71.4	854	1	QRHULD	LDL receptor precu
35	30	71.4	862	1	QRHULD	LDL receptor precu
36	30	71.4	871	2	AE0270	DNA topoisomerase
37	30	71.4	890	2	B85530	initiation factor
38	30	71.4	890	2	B81576	translation initia
39	30	71.4	890	2	F72093	H+-transporting At
40	30	71.4	949	2	T02083	hypothetical prote
41	30	71.4	1216	2	H85023	tripeptidyl-peptid
42	30	71.4	1354	2	T13930	prophage p13 prote
43	29	69.0	131	2	G86798	hypothetical prote
44	29	69.0	135	2	D71659	ribosomal protein
45	29	69.0	230	2	S73229	

## ALIGNMENTS

## RESULT 1

T42584

tegument protein 40 - equine herpesvirus 4 (strain NS80567)

C:Species: equine herpesvirus 4

A:Variety: strain NS80567

C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000

C:Accession: T42584

R:Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.

J. Gen. Virol. 79, 1197-1203, 1998

A:Title: The DNA sequence of equine herpesvirus-4.

A:Reference number: Z22173; MUID:98264497

A:Accession: T42584

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-529 <TEL>

A:Cross-references: EMBL:AF030027; NID:92605950; PIDN:AC59558.1; PID:92605966

A:Experimental source: strain NS80567

C:Genetics:

A:Gene: 40

C:Superfamily: varicella-zoster virus gene 38 protein

Query Match 81.0%; Score 34; DB 2; Length 529;

Best Local Similarity 87.5%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVNLLAEF 9

DB 423 EANLLAEF 430

## RESULT 2

B96999

methy1-accepting chemotaxis protein [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: B96999

R:Rolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B96999

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-668 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78781.1; PID:g15023694; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0805

Query Match 78.6%; Score 33; DB 2; Length 668;

Best Local Similarity 75.0%; Pred. No. 39;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLAE 8  
 ||:||||  
 Db 71 SEINLAE 78

RESULT 3  
 ORHULD  
 LDL receptor precursor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 17-May-1995 #sequence\_revision 17-May-1985 #text\_change 22-Jun-1999  
 C:Accession: A01383; A44176; A46751; S66642  
 R:Yamamoto, T.; Davis, C.G.; Brown, M.S.; Schneider, W.J.; Casey, M.L.; Goldstein, J.L.; Cell 39, 27-38, 1984  
 A:Title: The human LDL receptor: a cysteine-rich protein with Alu sequences in it its mRNA  
 A:Reference number: A01383; MUID:85024898  
 A:Accession: A01383  
 A:Molecule type: mRNA  
 A:Residues: 1-860 <YAM>  
 A:Cross-references: GB:L00352; GB:K02573; NID:g187094; PIDN:AAA56833.1; PID:g307121  
 R:Stedhof, T.C.; Goldstein, J.L.; Brown, M.S.; Russell, D.W.  
 Science 228, 815-822, 1985  
 A:Title: The LDL receptor gene: a mosaic of exons shared with different proteins.  
 A:Reference number: A44176; MUID:85218750  
 A:Accession: A44176  
 A:Molecule type: DNA  
 A:Status: nucleic acid sequence not shown  
 A:Residues: 1-184;193-314 <SUE>  
 A:Cross-references: GB:M10664  
 A>Note: translation of the nucleotide sequence is not complete  
 R:Russell, D.W.  
 submitted to GenBank, March 1985  
 A:Reference number: A46751  
 A:Accession: A46751  
 A:Molecule type: DNA  
 A:Residues: 1-860 <RUS>  
 A:Cross-references: GB:L00352; GB:K02573; NID:g187094; PIDN:AAA56833.1; PID:g307121  
 R:Yokode, M.; Pathak, R.K.; Hammer, R.E.; Brown, M.S.; Goldstein, J.L.; Anderson, R.G.W. J. Cell Biol. 117, 39-46, 1992  
 A:Title: Cytoplasmic sequence required for basolateral targeting of LDL receptor in live  
 A:Reference number: A44535; MUID:92210646  
 A:Contents: annotation: basolateral targeting signal  
 R:Chen, W.J.; Goldstein, J.L.; Brown, M.S.  
 J. Biol. Chem. 265, 3116-3123, 1990  
 A:Title: NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-  
 A:Reference number: A44541; MUID:90153958  
 A:Contents: annotation: internalization signal  
 R:Bieri, S.; Djordjevic, J.T.; Jamshidi, N.; Smith, R.; Kroon, P.A.  
 PNAS Lett. 371, 341-344, 1995  
 A:Title: Expression and disulfide-bond connectivity of the second ligand-binding repeat  
 A:Reference number: S66642; MUID:96013166  
 A:Contents: annotation: disulfide bonds  
 C:Comment: This transmembrane glycoprotein binds LDL, the major cholesterol-carrying lip  
 r-ligand complexes must first cluster into clathrin-coated pits.  
 C:Comment: An intrastand recombination event between two Alu sequences in the 3' untran  
 and cytoplasmic domains. Most of the receptors produced are secreted, but those that ad  
 and complexes are not internalized.  
 C:Genetics:  
 A:Gene: GDB:LDLR  
 A:Cross-references: GDB:119362; OMIM:143890  
 A:Map position: 19p13.2-19p13.2  
 A:Introns: 22/1: 64/1: 105/1: 23/1: 273/1: 314/1: 354/1: 396/1: 453/2: 529/2: 569/1: 61  
 C:Superfamily: LDL receptor; EGF homology; LDL receptor ligand-binding repeat homology;  
 C:Keywords: cholesterol; coated pits; duplication; endocytosis; glycoprotein; LDL; lipid  
 F:1-21/Domain: signal sequence #status predicted <Sig>  
 F:22-60/Product: LDL receptor #status predicted <EXT>  
 F:22-788/Domain: extracellular #status predicted <EXT>  
 F:27-63/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F:68-104/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:109-143/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:148-184/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F:197-231/Domain: LDL receptor ligand-binding repeat homology <LDL5>

F:236-270/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
 F:276-313/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
 F:318-352/Domain: EGF homology <EG1>  
 F:358-392/Domain: EGF homology <EG2>  
 F:399-438/Domain: LDL receptor WTD-containing repeat homology <YW1>  
 F:439-485/Domain: LDL receptor WTD-containing repeat homology <YW2>  
 F:486-528/Domain: LDL receptor WTD-containing repeat homology <YW3>  
 F:529-572/Domain: LDL receptor WTD-containing repeat homology <YW4>  
 F:573-615/Domain: LDL receptor WTD-containing repeat homology <YW5>  
 F:616-658/Domain: LDL receptor WTD-containing repeat homology <YW6>  
 F:667-711/Domain: EGF homology <EG3>  
 F:721-768/Region: clustered O-linked oligosaccharides  
 F:789-810/Domain: transmembrane #status predicted <TM>  
 F:811-860/Domain: intracellular #status predicted <INT>  
 F:825-829/Region: coated-pit mediated internalization signal  
 F:837-849/Region: basolateral targeting signal  
 F:86-82/75-95,89-104/Disulfide bonds: #status experimental  
 F:97,156,272,515,657/Binding site: carbonylate (asn) (covalent) #status predicted  
 F:318-329,325-338,340-352,358-368,364-377,379-392,667-681,677-696,698-711/Disulfide b

Query Match 78.6%; Score 33; DB 1; Length 860;  
 Best Local Similarity 87.5%; Pred. No. 52;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLAE 8  
 ||:|||||  
 Db 637 SDVNLAE 644

RESULT 4  
 G72351  
 conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: G72351  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
 C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: G72351  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-132 <ARN>  
 A:Cross-references: GB:AE001738; GB:AE000512; NID:g4981158; PIDN:AD35732.1; PID:g498  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0648

Query Match 76.2%; Score 32; DB 2; Length 132;  
 Best Local Similarity 66.7%; Pred. No. 10;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLAE 9  
 ||:|||||  
 Db 86 SEINLAE 94

RESULT 5  
 T10822  
 1-aminocyclopropane-1-carboxylate synthase (BC 4.4.1.14) - kidney bean (fragment)  
 N:Alternate names: ACC synthase  
 C:Species: Phaseolus vulgaris (kidney bean)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 04-Feb-2000  
 C:Accession: T10822  
 R:Pidgen, C.M.; Reid, D.M.  
 submitted to the EMBL Data Library, March 1998  
 A:Description: The role of light in controlling ACC synthase activity and gene expres  
 A:Reference number: Z17175  
 A:Accession: T10822  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-279 <PID>  
 A:Cross-references: EMBL:AF053355; NID:g2995852; PID:g2995853  
 A:Experimental source: cultivar Taylor, leaf  
 C:Genetics:  
 A:Note: ACS1  
 A:Note: Intron positions not resolved (incomplete sequence)  
 C:Function:  
 A:Description: catalyzes the formation of 1-aminocyclopropane-1-carboxylic acid from S-adenosyl-L-methionine  
 A:Pathway: ethylene biosynthesis  
 C:Superfamily: 1-aminocyclopropane-1-carboxylate synthase  
 C:Keywords: carbon-sulfur lyase; ethylene biosynthesis; phosphoprotein; pyridoxal phosphate  
 F:10/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match  
 Best Local Similarity 76.2%; Score 32; DB 2; Length 279;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 |||||:|  
 Db 129 SEVNLVDF 137

RESULT 6  
 T09120  
 basic leucine zipper protein - spinach  
 C:Species: Spinacia oleracea (spinach)  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 20-Jun-2000  
 C:Accession: T09120  
 R:Boile, C.; Lueberstedt, T.; Herranen, M.; Herrmann, R.; Oelmüller, R.  
 submitted to the EMBL Data Library, January 1998  
 A:Description: Molecular characterization of the spinach G-box binding protein family.  
 A:Reference number: Z16572  
 A:Accession: T09120  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-422 <BOL>  
 A:Cross-references: EMBL:AJ223624  
 C:Genetics:  
 A:Gene: bZIP  
 C:Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology

Query Match  
 Best Local Similarity 76.2%; Score 32; DB 2; Length 422;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAE 8  
 |||||:|  
 Db 324 SEVNLVLE 331

RESULT 7  
 T10854  
 1-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14) 7 - mung bean  
 N:Alternate names: ACC synthase  
 C:Species: Vigna radiata (mung bean)  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 04-Feb-2000  
 C:Accession: T10854  
 R:Kim, W.  
 submitted to the EMBL Data Library, August 1995  
 A:Reference number: Z17182  
 A:Accession: T10854  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-467 <KIM>  
 A:Cross-references: EMBL:U34987; NID:g1006806; PID:g1006807  
 A:Experimental source: tissue-type etiolated hypocotyl  
 C:Genetics:  
 A:Gene: ACS7  
 C:Function:  
 A:Description: catalyzes the rate-limiting first step of ethylene biosynthesis, the formation of S-adenosyl-L-methionine  
 A:Pathway: ethylene biosynthesis  
 C:Superfamily: 1-aminocyclopropane-1-carboxylate synthase  
 C:Keywords: carbon-sulfur lyase; ethylene biosynthesis; phosphoprotein; pyridoxal phosphate

F:272/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match  
 Best Local Similarity 76.2%; Score 32; DB 2; Length 467;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 |||||:|  
 Db 211 SEVNLVDF 219

RESULT 8  
 T10889  
 1-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14) 6 - mung bean  
 N:Alternate names: ACC synthase; S-adenosyl-L-methionine methyltransferase  
 C:Species: Vigna radiata (mung bean)  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
 C:Accession: T10889; T10842  
 R:Yoon, I.; Mori, H.; Kim, J.; Kang, B.; Imaseki, H.  
 Plant Cell Physiol. 38, 217-224, 1997  
 A:Title: VR-ACS6 is an auxin-inducible 1-aminocyclopropane-1-carboxylate synthase  
 A:Reference number: Z17198; MUID:97294927  
 A:Accession: T10889  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-472 <YOO>  
 A:Cross-references: EMBL:AB000679; NID:g1813330; PIDN:BAJ19161.1; PID:g1813331  
 R:Kim, W.  
 submitted to the EMBL Data Library, August 1995  
 A:Reference number: Z17182  
 A:Accession: T10842  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-37, 'Q', '39-124', 'E', '126-389', 'QE', '392-425', 'NG', '428-442', 'T', '444-472' <KIM>  
 A:Cross-references: EMBL:U34986; NID:g1006804; PID:g1006805  
 A:Experimental source: tissue-type etiolated hypocotyls  
 C:Genetics:  
 A:Gene: ACS6  
 C:Function:  
 A:Description: catalyzes the conversion of S-adenosylmethionine to 1-aminocyclopropane-1-carboxylate  
 A:Pathway: ethylene biosynthesis  
 C:Superfamily: 1-aminocyclopropane-1-carboxylate synthase  
 C:Keywords: carbon-sulfur lyase; ethylene biosynthesis; phosphoprotein; pyridoxal phosphate  
 F:275/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match  
 Best Local Similarity 76.2%; Score 32; DB 1; Length 472;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 |||||:|  
 Db 211 SEVNLVDF 219

RESULT 9  
 D95318  
 N1FE oxidoreductase nife [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid  
 C:Species: Sinorhizobium meliloti  
 C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: D95318  
 R:Barneit, M.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B.  
 Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yen, K.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: D95318  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-476 <KUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK65110.1; PID:g14523548; GSPDB:GN00165  
 A:Experimental source: strain 1021, megaplasmid pSymA  
 R:Gilbert, F.; Finn, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, F.;  
 Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.

L.: Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandemol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The complete genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: nifE  
 A:Genome: Plasmid  
 C:Superfamily: dinitrogenase alpha chain; nitrogenase vanadium-iron protein alpha chain

Query Match 76.2%; Score 32; DB 2; Length 476;  
 Best Local Similarity 55.6%; Pred. No. 44;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 Db 195 SDINLGEF 203

RESULT 10  
 E82333  
 oligoribonuclease VC0341 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: E82333  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.  
 L.R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A:Reference number: A82035; MUID:20406833

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-181 <HE>  
 A:Cross-references: GB:AE004123; GB:AE003852; NID:99654756; PIDN:AAF93514.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC0341  
 C:Map position: 1  
 C:Superfamily: human conserved hypothetical protein DKF2P566E144.1

Query Match 73.8%; Score 31; DB 2; Length 181;  
 Best Local Similarity 75.0%; Pred. No. 25;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLAEF 8  
 Db 35 SEVNLAEF 42

RESULT 11  
 T17795  
 hypothetical protein A298L - Chlorella virus PCV-1  
 C:Species: Chlorella virus PCV-1  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T17795  
 R:Graves, M.V.; Van Etten, J.L.  
 submitted to the EMBL Data Library, May 1999  
 A:Reference number: Z18806

A:Accession: T17795  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-225 <GRA>  
 A:Cross-references: EMBL:U42580; NID:94028896; PIDN:AA09666.1  
 A:Experimental source: specific host *Chlorella* strain NC64A  
 C:Genetics:  
 A:Note: A298L

Query Match 73.8%; Score 31; DB 2; Length 225;  
 Best Local Similarity 85.7%; Pred. No. 32;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 VNLAEF 9  
 Db 9 VNLAEF 15

RESULT 12

AD2820  
 conserved hypothetical protein Atu1983 [imported] - Agrobacterium tumefaciens (strain  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: AD2820  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
 erge, G.; Gillet, W.; Grant, C.; Genthner, D.; Kutayav, T.; Levy, R.; Li, M.; McCl  
 : Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Ioo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
 ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AD2820  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-236 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AAL42978.1; PID:g17740438; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu1983  
 A:Map position: circular chromosome

Query Match 73.8%; Score 31; DB 2; Length 236;  
 Best Local Similarity 66.7%; Pred. No. 34;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 Db 137 SQVNLAASY 145

RESULT 13

D97598  
 hypothetical protein AGR\_C\_3608 [imported] - Agrobacterium tumefaciens (strain C58, C  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
 C:Accession: D97598  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldm  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,  
 Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*  
 A:Reference number: A97359; PMID:11743194

A:Accession: D97598  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-240 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK8741.1; PID:g15157106; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_3608  
 A:Map position: circular chromosome

Query Match 73.8%; Score 31; DB 2; Length 240;  
 Best Local Similarity 66.7%; Pred. No. 34;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
 Db 141 SQVNLAASY 149

RESULT 14

H87588  
 hypothetical protein CC2743 [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus



C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C:Accession: H87588  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A:Reference number: AB7249; MUID:21173698; PMID:11259647  
 A:Accession: H87588  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-394 <STO>  
 A:Cross-references: GB:AE005673; NID:913424336; PIDN:AAK24708.1; GSPDB:GN00148  
 C:Genetics:  
 A:Gene: CC2743

Query Match 73.8%; Score 31; DB 2; Length 394;  
 Best Local Similarity 87.5%; Pred. No. 60;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLAE 8  
 |||||  
 Db 370 SEVNLXAE 377

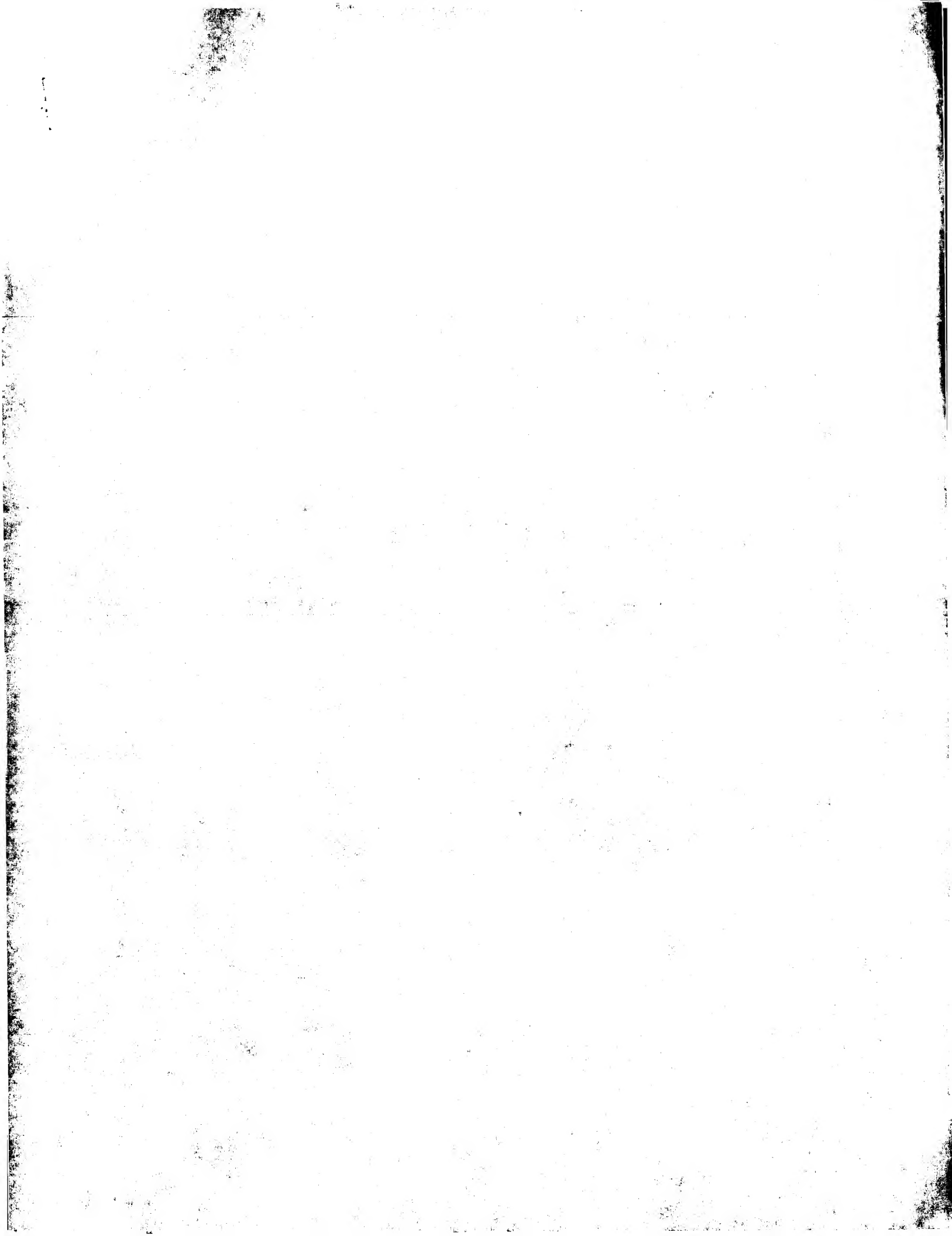
# RESULT 15

T15020  
 probable porphyrin biosynthetic enzyme - *Yersinia pestis* plasmid PMT1  
 C:Species: *Yersinia pestis*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Sep-1999  
 C:Accession: T15020; T14700  
 R:Jandler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.  
 Infect. Immun. 66, 5731-5742, 1998  
 A:Title: Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid  
 A:Reference number: Z18268; MUID:99043898  
 A:Accession: T15020  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-411 <LIN>  
 A:Cross-references: EMBL:AF074611; NID:93883003; PID:93883103; PIDN:AAC82763.1  
 R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Garnea, J.; Kobayashi, A.; Carrano,  
 submitted to the EMBL Data Library, March 1998  
 A:Description: Structural organization of virulence determinants in three *Yersinia pestis*  
 A:Reference number: Z18168  
 A:Accession: T14700  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-411 <HUP>  
 A:Cross-references: EMBL:AF053947; NID:92996286; PID:92996336; PIDN:AAC13216.1  
 C:Genetics:  
 A:Gene: COB5  
 A:Genome: plasmid PMT1

Query Match 73.8%; Score 31; DB 2; Length 411;  
 Best Local Similarity 75.0%; Pred. No. 63;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEVNLAE 8  
 :|||:|  
 Db 326 AEVNLAE 333

Search completed: October 30, 2002, 12:31:37  
 Job time : 4.72482 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 ; Search time 1.28256 Seconds  
(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-91

Perfect score: 42

Sequence: 1 SEVNLLAEF 9

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PCRTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfilest1.pep: \*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	81.0	9	3	US-08-802-981-219
2	34	81.0	10	2	US-08-659-984A-19
3	34	81.0	10	4	US-08-660-531-19
4	34	81.0	11	5	PCT-US94-07043A-3
5	34	81.0	21	2	US-08-659-984A-18
6	34	81.0	21	3	US-08-802-981-112
7	34	81.0	21	4	US-08-660-531-18
8	34	81.0	30	2	US-08-659-984A-17
9	34	81.0	30	4	US-08-660-531-17
10	34	81.0	33	2	US-08-659-984A-16
11	34	81.0	33	4	US-08-660-531-16
12	34	81.0	42	2	US-08-659-984A-15
13	34	81.0	42	4	US-08-660-531-15
14	34	81.0	506	2	US-08-659-984A-21
15	34	81.0	506	4	US-08-660-531-21
16	34	81.0	506	4	US-09-054-334-4
17	33	78.6	280	4	US-08-652-877-6
18	33	78.6	860	4	US-08-476-515A-6
19	33	78.6	860	1	US-08-092-817-4
20	33	78.6	1410	2	US-08-470-058-4
21	33	78.6	1410	3	US-09-037-188-4
22	33	78.6	1410	4	US-09-285-310-4
23	32	76.2	942	4	US-09-171-461-12
24	30	71.4	40	4	US-07-861-458C-118
25	30	71.4	67	6	5196333-10
26	30	71.4	103	3	US-08-339-708A-12
27	30	71.4	493	6	5196333-4

28	30	71.4	564	4	US-09-360-197-16	Sequence 16, Appl
29	30	71.4	753	4	US-07-861-458C-98	Sequence 98, Appl
30	30	71.4	753	4	US-07-861-458C-99	Sequence 99, Appl
31	29	69.0	27	1	US-08-141-324-12	Sequence 12, Appl
32	29	69.0	27	1	US-08-541-902-12	Sequence 12, Appl
33	29	69.0	27	1	US-08-942-012B-33	Sequence 33, Appl
34	28	66.7	9	3	US-08-802-981-220	Sequence 220, App
35	28	66.7	9	3	US-08-802-981-223	Sequence 223, App
36	28	66.7	21	3	US-08-802-981-113	Sequence 113, App
37	28	66.7	21	3	US-08-802-981-116	Sequence 116, App
38	28	66.7	74	1	US-08-543-238-5	Sequence 5, Appl
39	28	66.7	74	1	US-08-420-528-5	Sequence 5, Appl
40	28	66.7	237	2	US-08-933-750C-32	Sequence 32, Appl
41	28	66.7	237	4	US-09-234-613-32	Sequence 32, Appl
42	27	64.3	12	4	US-09-399-494-22	Sequence 22, Appl
43	27	64.3	31	4	US-09-399-494-11	Sequence 11, Appl
44	27	64.3	34	3	US-09-044-536A-21	Sequence 21, Appl
45	27	64.3	64	2	US-08-687-559-3	Sequence 3, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-802-981-219
; Sequence 219, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-000300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 219:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-802-981-219
```

Query Match 81.0%; Score 34; DB 3; Length 9;  
Best Local Similarity 88.9%; Pred. No. 1.7e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 SEVNLLAEF 9  
Db 1 SEVNLLAEF 9

RESULT 2  
US-08-659-984A-19  
Sequence 19, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-0028100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-659-984A-19

Query Match 81.0%; Score 34; DB 2; Length 10;  
Best Local Similarity 88.9%; Pred. No. 0.33;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9  
DB 1 SEVNLDAEF 9

RESULT 3  
US-08-660-531-19  
Sequence 19, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-0022100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-660-531-19

Query Match 81.0%; Score 34; DB 4; Length 10;  
Best Local Similarity 88.9%; Pred. No. 0.33;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9  
DB 1 SEVNLDAEF 9

RESULT 4  
PCT-US94-07043A-3  
Sequence 3, Application PC/TUS9407043A  
GENERAL INFORMATION:  
APPLICANT: Tamburini, Paul P.; Benz, G nter; H bich,  
APPLICANT: Dieter; Dreyer, Robert N.; Koenig, Gerhard  
TITLE OF INVENTION: CATHEPSIN D IS AN AMYLOIDGENIC  
TITLE OF INVENTION: PROTEASE IN ALZHEIMER S DISEASE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESS: Miles Inc.  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06516  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: Sharp PC 4600  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07043A  
FILING DATE: June 21, 1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10889

FILING DATE: November 12, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/995,660  
FILING DATE: December 16, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/880,914  
FILING DATE: May 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pamela A. Simonton  
REGISTRATION NUMBER: 31,060  
REFERENCE/DOCKET NUMBER: MTI 224.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 937-2340  
TELEFAX: (203) 937-2795  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US94-07043A-3

Query Match 81.0%; Score 34; DB 5; Length 11;  
Best Local Similarity 88.9%; Pred. No. 0.36;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAER 9  
|||||  
DB 2 SEVNLAER 10

RESULT 5  
US-08-659-984A-18  
Sequence 18, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Simha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-659-984A-18

Query Match 81.0%; Score 34; DB 2; Length 21;  
Best Local Similarity 88.9%; Pred. No. 0.74;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAER 9  
|||||  
DB 1 SEVNLAER 9

RESULT 6  
US-08-802-981-112  
Sequence 112, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product="Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product="Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /product="Acp"  
US-08-802-981-112

Query Match 81.0%; Score 34; DB 3; Length 21;  
Best Local Similarity 88.9%; Pred. No. 0.74;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAER 9  
|||||  
DB 6 SEVNLAER 14

RESULT 7  
US-08-660-531-18  
; Sequence 18, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S. P.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002210US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-660-531-18  
Query Match 81.0%; Score 34; DB 4; Length 21;  
Best Local Similarity 88.9%; Pred. No. 0.74;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNDAER 9  
Db 1 SEVNDAER 9

RESULT 8  
US-08-659-984A-17  
; Sequence 17, Application US/08659984A  
; Patent No. 5942400  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Jacobson-Croak, Kirsten L.  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California

COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-17  
Query Match 81.0%; Score 34; DB 2; Length 30;  
Best Local Similarity 88.9%; Pred. No. 1.1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNDAER 9  
Db 22 SEVNDAER 30

RESULT 9  
US-08-660-531-17  
; Sequence 17, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002210US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-17

Query Match 81.0%; Score 34; DB 4; Length 30;  
Best Local Similarity 88.9%; Pred. No. 1.1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLAEF 9  
|||||  
Db 22 SEVNLAEF 30

RESULT 10  
US-08-659-984A-16  
Sequence 16, Application US/08659984A  
Patent No. 5842400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-16

Query Match 81.0%; Score 34; DB 2; Length 33;  
Best Local Similarity 88.9%; Pred. No. 1.2;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLAEF 9  
|||||  
Db 22 SEVNLAEF 30

Db 13 SEVNLAEF 21

RESULT 11  
US-08-660-531-16  
Sequence 16, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-16

Query Match 81.0%; Score 34; DB 4; Length 33;  
Best Local Similarity 88.9%; Pred. No. 1.2;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SEVNLAEF 9  
|||||  
Db 13 SEVNLAEF 21

RESULT 12  
US-08-659-984A-15  
Sequence 15, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco

STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-15

Query Match 81.0%; Score 34; DB 2; Length 42;  
Best Local Similarity 88.9%; Pred. No. 1.6;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLAEF 9  
|||||  
Db 22 SEVNLDAEF 30

RESULT 13  
US-08-660-531-15  
Sequence 15, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541

REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match 81.0%; Score 34; DB 4; Length 42;  
Best Local Similarity 88.9%; Pred. No. 1.6;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLAEF 9  
|||||  
Db 22 SEVNLDAEF 30

RESULT 14  
US-08-659-984A-21  
Sequence 21, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-659-984A-21

Query Match 81.0%; Score 34; DB 2; Length 506;  
Best Local Similarity 88.9%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLAEF 9  
|||||



Db 403 SEVNLDAEF 411

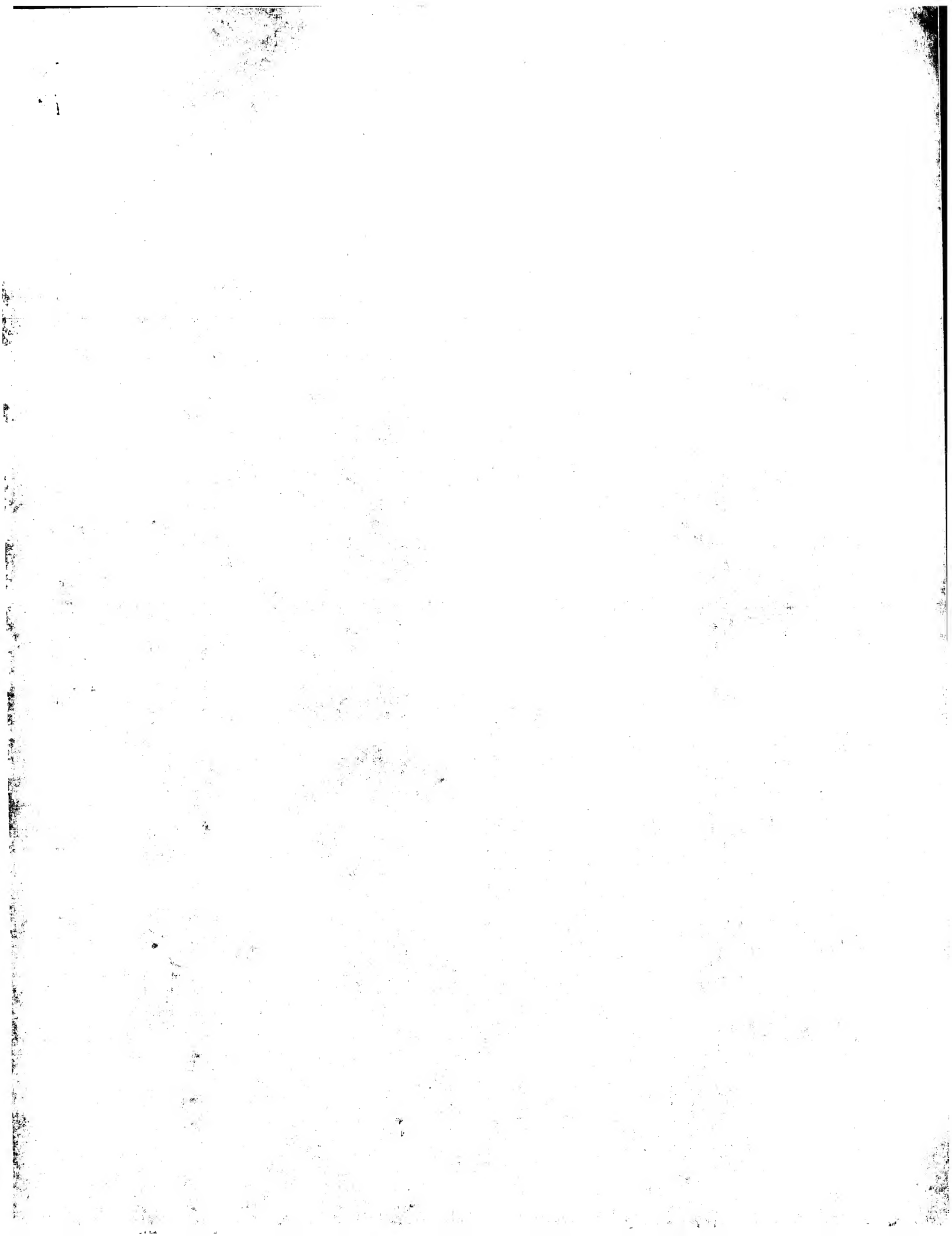
RESULT 15  
US-08-660-531-21  
; Sequence 21, Application US/08660531  
; Patent NO. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Kelm, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002210US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 506 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-660-531-21

Query Match 81.0%; Score 34; DB 4; Length 506;  
Best Local Similarity 88.9%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEVNLDAEF 9  
|||||  
Db 403 SEVNLDAEF 411

Search completed: October 30, 2002, 12:32:35  
Job time : 2.28256 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:21:30 ; Search time 2.49631 Seconds  
(without alignments)  
554.401 Million cell updates/sec

Title: US-09-724-571-81

Perfect score: 34  
Sequence: 1 EVXVAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	29	85.3	195	16	097F65	097F65	clostridium
2	28	82.4	776	10	094IL7	094IL7	populus tre
3	28	82.4	3414	12	091C40	091C40	langat viru
4	28	82.4	3414	12	091G39	091G39	langat viru
5	27	79.4	184	17	026364	026364	methanother
6	27	79.4	201	16	034914	034914	bacillus su
7	27	79.4	338	17	09HL69	09HL69	thermoplasm
8	27	79.4	598	17	058774	058774	pyrococcus
9	27	79.4	601	16	092U63	092U63	thiobium m
10	27	79.4	775	10	09FE38	09FE38	arabidopsis
11	27	79.4	802	10	09SUR2	09SUR2	arabidopsis
12	27	79.4	1158	5	096554	096554	caenorhabdi
13	27	79.4	1560	5	026644	026644	strongyloce
14	26	76.5	46	16	092ZL0	092ZL0	rhizobium m
15	26	76.5	92	9	09FZT5	09FZT5	pseudomonas
16	26	76.5	150	2	052322	052322	escherichia

17	26	76.5	194	10	0946P3	0946P3	uncultured
18	26	76.5	252	16	0927W5	0927W5	chlamydia p
19	26	76.5	278	10	09FXF4	09FXF4	arabidopsis
20	26	76.5	294	10	09M2H4	09M2H4	arabidopsis
21	26	76.5	324	10	092P05	092P05	arabidopsis
22	26	76.5	363	4	09H266	09H266	homo sapien
23	26	76.5	490	16	092T24	092T24	rhizobium m
24	26	76.5	498	16	09KN11	09KN11	vibrio chol
25	26	76.5	533	16	092BS4	092BS4	listeria in
26	26	76.5	543	16	09KDI2	09KDI2	bacillus ha
27	26	76.5	559	1	09UWN9	09UWN9	sulfolobus
28	26	76.5	559	1	055088	055088	sulfolobus
29	26	76.5	561	17	095867	095867	sulfolobus
30	26	76.5	562	17	09U210	09U210	pyrococcus
31	26	76.5	579	2	09K2D3	09K2D3	strepomyce
32	26	76.5	587	16	09U2P8	09U2P8	neisseria m
33	26	76.5	587	16	09U0T3	09U0T3	neisseria m
34	26	76.5	623	16	09PAA1	09PAA1	xyella fas
35	26	76.5	642	3	09C260	09C260	neurospora
36	26	76.5	778	11	099MP4	099MP4	mus musculu
37	26	76.5	791	2	09P2E2	09P2E2	homo sapien
38	26	76.5	821	2	09XBW4	09XBW4	porphyromon
39	26	76.5	1014	16	09K9C6	09K9C6	bacillus ha
40	26	76.5	1085	5	001585	001585	caenorhabdi
41	26	76.5	1249	5	097043	097043	drosophila
42	26	76.5	1249	5	09VF33	09VF33	drosofila
43	25	73.5	127	12	091BF3	091BF3	spodoptera
44	25	73.5	136	11	063172	063172	rattus norv
45	25	73.5	137	4	09N248	09N248	homo sapien

## ALIGNMENTS

RESULT 1

097F65 ID 097F65 PRELIMINARY: PRT: 195 AA.

AC 097F65: 01-OCT-2001 (TREMBLrel. 18. Created)

DT 01-OCT-2001 (TREMBLrel. 18. Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19. Last annotation update)

DE THYMIDINE KINASE.

GN CAC2887.

OS Clostridium acetobutylicum.

OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

CC Clostridium.

OK NCBI\_TaxID=1488;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RX MEDLINE=21359325; PubMed=11466286;

RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I., Tatusov R.L., Sabathe F., Doucet-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;

RT \*Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum";

RL J. Bacteriol. 183:4823-4838(2001).

DR EMBL: AE007786; AK80830.1; -

DR InterPro: IPR000345; CytC\_heme\_bind.

DR InterPro: IPR001267; Thymd\_kin\_cell.

DR Pfam: PF00265; TK; 1.

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN\_1.

DR PROSITE: PS00603; TK\_CELLULAR\_TYPE; UNKNOWN\_1.

KW kinase; Complete proteome.

SQ SEQUENCE 195 AA; 22303 MW; 6EDB551986B203BA CRC64;

Query Match 85.3%; Score 29; DB 16; Length 195;

Best Local Similarity 62.5%; Pred. No. 18;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVXVAEF 8

|||||

Db 130 ELMATAEF 137

## RESULT 2

0941L7 PRELIMINARY; PRT: 776 AA.

AC 0941L7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE HIGH-AFFINITY POTASSIUM UPTAKE TRANSPORTER.  
 GN KUP1.  
 OS Populus tremula x Populus tremuloides.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids I; Malpighiales; Salicaceae; Populus.  
 OX NCBI\_TaxID=47664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Langer K., Ache P., Fromm J., Hedrich R.;  
 RT "PKUP contributes to wood formation.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ299422; CAC39168.1;  
 SQ SEQUENCE 776 AA; 87303 MW; BEC03D57ED0869BC CRC64;

Query Match 82.4%; Score 28; DB 10; Length 776;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 VMXVAEF 8  
 |||||  
 Db 604 VMSVAEF 610

## RESULT 3

091G40 PRELIMINARY; PRT: 3414 AA.

AC 091G40;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE POLYPROTEIN PRECURSOR.  
 OS Langat virus (strain TP21).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxID=31638;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20192178; PubMed=10725214;  
 RX Campbell M.S., Pietnev A.G.;  
 RT "Infectious cDNA clones of Langat tick-borne flavivirus that differ  
 from their parent in peripheral neurovirulence.";  
 RT Virology 269:225-237(2000).  
 RL EMBL: AF253419; AAF75259.1;  
 DR HSSP: P14336; 1SVB.  
 DR MEROPS: S07.001;  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR000336; Flavi\_glycoprote.  
 DR InterPro: IPR001850; Flavi\_helicase.  
 DR InterPro: IPR001157; Flavi\_NSI.  
 DR InterPro: IPR001528; Flavi\_NS4B.  
 DR InterPro: IPR000208; Flavi\_NS5.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00869; Flavi\_glycoprote; 1.  
 DR Pfam: PF00949; Flavi\_helicase; 1.  
 DR Pfam: PF00948; Flavi\_NSI; 1.  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR ProDom: PD001496; Flavi\_NSI; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR SMART; SM00490; HELIC; 1.  
 KW Signal.

FT SIGNAL 97 116 POTENTIAL.  
 FT CHAIN 1 96 CAPSID PROTEIN.  
 FT CHAIN 117 280 MEMBRANE-ASSOCIATED GLYCOPROTEIN  
 FT CHAIN 281 779 PRECURSOR  
 FT CHAIN 780 1128 ENVELOPE MEMBRANE-ASSOCIATED  
 FT CHAIN 1129 1358 GLYCOPROTEIN.  
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 2111 2239 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS3.  
 FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS4.  
 FT CHAIN 3414 378135 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 3414 8DE86A46A0E8F7E3 CRC64;  
 SQ SEQUENCE 3414 AA; 378135 MW; 8DE86A46A0E8F7E3 CRC64;

Query Match 82.4%; Score 28; DB 12; Length 3414;  
 Best Local Similarity 85.7%; Pred. No. 5.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 VMXVAEF 8  
 |||||  
 Db 929 VMTVAEF 935

## RESULT 4

091G39 PRELIMINARY; PRT: 3414 AA.

AC 091G39;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE POLYPROTEIN PRECURSOR.  
 OS Langat virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxID=11085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20192178; PubMed=10725214;  
 RX Campbell M.S., Pietnev A.G.;  
 RT "Infectious cDNA clones of Langat tick-borne flavivirus that differ  
 from their parent in peripheral neurovirulence.";  
 RT Virology 269:225-237(2000).  
 RL EMBL: AF253420; AAF75260.1;  
 DR HSSP: P14336; 1SVB.  
 DR MEROPS: S07.001;  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR000336; Flavi\_glycoprote.  
 DR InterPro: IPR001850; Flavi\_helicase.  
 DR InterPro: IPR001157; Flavi\_NSI.  
 DR InterPro: IPR001528; Flavi\_NS4B.  
 DR InterPro: IPR000208; Flavi\_NS5.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00869; Flavi\_glycoprote; 1.  
 DR Pfam: PF00949; Flavi\_helicase; 1.  
 DR Pfam: PF00948; Flavi\_NSI; 1.  
 DR Pfam: PF01349; Flavi\_NS4B; 1.  
 DR Pfam: PF00972; Flavi\_NS5; 1.  
 DR ProDom: PD001496; Flavi\_NSI; 1.  
 DR ProDom: PD001556; Flavi\_glycoprote; 1.  
 DR SMART; SM00490; HELIC; 1.  
 KW Signal.  
 FT SIGNAL 97 116 POTENTIAL.  
 FT CHAIN 1 96 CAPSID PROTEIN.  
 FT CHAIN 117 280 MEMBRANE-ASSOCIATED GLYCOPROTEIN  
 FT CHAIN 281 779 PRECURSOR  
 FT CHAIN 780 1128 ENVELOPE MEMBRANE-ASSOCIATED  
 FT CHAIN 1129 1358 GLYCOPROTEIN.  
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 1490 2110 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 2110 2239 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS3.  
 FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS4.  
 FT CHAIN 3414 378135 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 3414 8DE86A46A0E8F7E3 CRC64;

```
FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2512 3414 NONSTRUCTURAL PROTEIN NS5.
SQ SEQUENCE 3414 AA; 378014 MW; BF8F8ACEAB96D534 CRC64;

Query Match
Best Local Similarity 82.4%; Score 28; DB 12; Length 3414;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 VMXVAEF 8
DB 929 VMXVAEF 935

RESULT 5
ID 026364 PRELIMINARY; PRT; 184 AA.
AC 026364;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JUN-1998 (TREMblrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
DE DNA-DEPENDENT RNA POLYMERASE, SUBUNIT E'.
GN MTH264.
OS Methanothermobacter thermautotrophicus.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicore R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pletrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nollinger J., Reeve J.N.;
RT *Complete genome sequence of Methanobacterium thermautotrophicum
RT delah: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7153(1997).
DR EMBL; AF000812; AAB84770.1; -.
DR HSSP; P05055; ISRO.
DR InterPro; IPR003029; S1.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 1.
KW Complete proteome.
SQ SEQUENCE 184 AA; 20695 MW; 37AAB7B39D631813 CRC64;

Query Match
Best Local Similarity 79.4%; Score 27; DB 17; Length 184;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
DB 88 EVMXVAEF 95

RESULT 6
ID 034914 PRELIMINARY; PRT; 201 AA.
AC 034914;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JUN-1998 (TREMblrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE YTHO.
GN YTHO.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=96312354; PubMed=8763940;
RA Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahim P., Pero J.;
RT "Cloning, sequencing, and characterization of the Bacillus subtilis
RT bclA biosynthetic operon.";
RL J. Bacteriol. 178:4122-4130(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich D.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriis R., Boutsier L., Brans A., Braun M., Briquell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.-J., Haga K., Halesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kodayashi Y., Koelter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nockack M.,
RA Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Pirescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadleir I.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Seliguchi J., Sekowska A., Serron S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terstira P., Togonou K.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT *The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF008220; AAC00267.1; -.
DR EMBL; Z99119; CAB14996.1; -.
KW Complete proteome.
SQ SEQUENCE 201 AA; 22733 MW; CBFAD9B6BA348F06 CRC64;

Query Match
Best Local Similarity 79.4%; Score 27; DB 16; Length 201;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVMXVAEF 8
DB 22 DIMDIAEF 29

RESULT 7
ID 09HL69 PRELIMINARY; PRT; 338 AA.
AC 09HL69;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN TAO362.
GN TAO362.
OS Thermoplasma acidophilum.
```

OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;  
 OC Thermoplasma.  
 OC NCBI\_TaxID=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 1728;  
 RX MEDLINE-20479972; PubMed-11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 acidophilum.";  
 RL Nature 407:508-513(2000).  
 DR EMBL; AL45064; CAC11506.1;  
 DR InterPro; IPR001005; Myb\_DNA\_bind.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 338 AA; 39348 MW; D86C95529E9A43E CRC64;

Query Match 79.4%; Score 27; DB 17; Length 338;  
 Best Local Similarity 62.5%; Pred. No. 1e+02; 1; Indels 0; Gaps 0;  
 Matches 5; Conservative 2; Mismatches 1;

OY 1 EVMYAEF 8  
 DB 329 EIMPLAEF 336

RESULT 8  
 ID 058774 PRELIMINARY; PRT; 598 AA.

AC 058774;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE HYPOTHEICAL 70.2 KDA PROTEIN PH1023.  
 GN PH1023.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OC NCBI\_TaxID=33953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OT3;  
 RX MEDLINE-98344137; PubMed-9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.I., Horikoshi K.,  
 RA Masuchi Y., Shizuza H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998)  
 DR EMBL; AF000004; BAA30120.1;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 598 AA; 70188 MW; 3E1ACD0F8EDC7661 CRC64;

Query Match 79.4%; Score 27; DB 17; Length 598;  
 Best Local Similarity 62.5%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMYAEF 8  
 DB 280 EIVMYAEF 287

RESULT 9  
 ID 092063 PRELIMINARY; PRT; 601 AA.

AC 092063;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PUTATIVE 1,4-ALPHA-GLOCAN BRANCHING ENZYME PROTEIN (EC

DE 2.4.1.18).  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSymB (megaplasmid 2).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OC NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RX MEDLINE-21396508; PubMed-11481431;  
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
 RA Golding B., Puehler A.;  
 RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliloti.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 DR EMBL; AL603646; CAC49676.1;  
 KW transferase; Glycosyltransferase; Plasmid; Hypothetical protein;  
 SO SEQUENCE 601 AA; 66954 MW; 024DA322A7B72C2E CRC64;

Query Match 79.4%; Score 27; DB 16; Length 601;  
 Best Local Similarity 62.5%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMYAEF 8  
 DB 151 EIMPEVQAF 158

RESULT 10  
 ID 09FE38 PRELIMINARY; PRT; 775 AA.

AC 09FE38;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TINY ROOT HAIR 1 PROTEIN.  
 GN TRH1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rigas S., Debrosses G., Haralampidis K., Vicente-Agullo F.,  
 RA Feldmann K.A., Grabov A., Dolan L., Hatzopoulos P.;  
 RT "A Potassium Transporter Required for Tip Growth in Arabidopsis.";  
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ296156; CAC16138.1;  
 DR EMBL; AJ296155; CAC16137.1;  
 DR InterPro; IPR003855; K+-transprtr.  
 DR Pfam; PF02705; K\_trans; 1.  
 SO SEQUENCE 775 AA; 86842 MW; B0C5068B48E8180 CRC64;

Query Match 79.4%; Score 27; DB 10; Length 775;  
 Best Local Similarity 71.4%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 VMXVAEF 8  
 DB 602 VMSIAEF 608

RESULT 11  
 ID 09SUR2 PRELIMINARY; PRT; 802 AA.

AC 09SUR2;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PUTATIVE POTASSIUM TRANSPORT PROTEIN.

GN F0016.110 OR ATG623640.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 RX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Medler H., Medler E., Mambutt R., Hohelsel J., Mewes H.W.,  
 RA Mayer K.F.X., Scheller C.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Medler H., Medler E., Mambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL033394; CAA23030.1; -;  
 DR EMBL; AL161559; CAB79319.1; -;  
 DR InterPro: IPR003855; K+\_transptr.  
 DR Pfam: PF02705; K\_trans; 1.  
 SO SEQUENCE 802 AA; 89817 MW; 617AF5F76B9BB60 CRC64;  
 Query Match 79.4%; Score 27; DB 10; Length 802;  
 Best Local Similarity 71.4%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 VMXVAEF 8  
 Db 629 VMSIAEF 635

RESULT 12  
 096554  
 ID 096554 PRELIMINARY; PRT; 1158 AA.  
 AC 096554;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN Y75B7AL.4.  
 GN Y75B7AL.4  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 RX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RC STRAIN-BRISTOL N2;  
 RA Lamar B.;  
 RL "The sequence of C. elegans cosmid Y75B7AL.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC024669; AAK68606.1; -;  
 SO SEQUENCE 1158 AA; 127146 MW; 63D522E9243FF01B CRC64;

Query Match 79.4%; Score 27; DB 5; Length 1158;  
 Best Local Similarity 75.0%; Pred. No. 3.5e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 EVXVAEF 8  
 Db 563 EOMXVAEF 570

RESULT 13  
 026644  
 ID 026644 PRELIMINARY; PRT; 1560 AA.  
 AC 026644;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CALCIUM-BINDING PROTEIN.  
 GN END016.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinozoa; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 RX NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Soltysik-Espanola M., Klinzing D.C., Pfarr K., Burke R.D., Ernst S.G.;  
 RL "Endo16, a large multi-domain protein found on the surface and ECM of  
 RL endodermal cells during sea urchin gastrulation, binds calcium.";  
 RL Dev. Biol. 0:0-0(1994).  
 DR EMBL; L34680; AAA30047.1; -;  
 DR InterPro: IPR001614; Myelin\_PLP.  
 DR Pfam: PF02014; Reeler.  
 DR SMART: SM00002; PLP; 1.  
 SO SEQUENCE 1560 AA; 175228 MW; 2D1A0E0D8E97CE70 CRC64;  
 Query Match 79.4%; Score 27; DB 5; Length 1560;  
 Best Local Similarity 62.5%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 EVXVAEF 8  
 Db 705 ELLVAEF 712

RESULT 14  
 0922L0  
 ID 0922L0 PRELIMINARY; PRT; 46 AA.  
 AC 0922L0;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN SMA0833.  
 GN SMA0833.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OG Plasmid pSymb (megaplasmid 1).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 RX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-1021;  
 RC MEDLINE=21396509; PubMed=11481432;  
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,  
 RA Barloy-Hubler F., Bowser L., Capela D., Gilbert F., Gouzy J.,  
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,  
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Teh K.-C., Davis R.W., Federspiel N.A., Long S.R.;  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Sinorhizobium meliloti pSymb megaplasmid.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 DR EMBL; AE007236; AAK65112.1; -;

KW plasmid: Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 46 AA; 5236 MW; 3AD9B653CB1B4813 CRC64;

Query Match 76.5%; Score 26; DB 16; Length 46;  
 Best Local Similarity 71.4%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAE 7  
 |::|::|  
 Db 17 ETMAVAE 23

## RESULT 15

09FZT5 PRELIMINARY; PRT; 92 AA.  
 AC 09FZT5;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)  
 DE P3B.  
 GN 3B.  
 OS Pseudomonas bacteriophage phi-13.  
 OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.  
 OX NCBI\_TaxID=134554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99350412; PubMed=10419946;  
 RA Mindich L., Qiao X., Qiao J., Onodera S., Romantschuk M.,  
 Hoogstraaten D.;  
 RT "Isolation of additional bacteriophages with genomes of segmented  
 RT double-stranded RNA."; *Bacteriol. 181:4505-4508(1999).*  
 RL J. Bacteriol. 181:4505-4508(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20473938; PubMed=11017801;  
 RA Qiao X., Qiao J., Onodera S., Mindich L.;  
 RT "Characterization of phi13, a bacteriophage related to phi6 and  
 RT containing three dsRNA genomic segments."; *Virology 275:218-224(2000).*  
 RL EMBL; AF261667; AAG00441.1; -;  
 DR EMBL; AF261667; AAG00441.1; -;  
 SQ SEQUENCE 92 AA; 9904 MW; B52E88EDB8720740 CRC64;

Query Match 76.5%; Score 26; DB 9; Length 92;  
 Best Local Similarity 50.0%; Pred. No. 51;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVMXVAEF 8  
 |::|::|  
 Db 47 EILSVSEF 54

Search completed: October 30, 2002, 12:30:10  
 Job time : 5.49631 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:20:50 ; Search time 3.60442 Seconds

(without alignments)  
277.344 Million cell updates/sec

Title: US-09-724-571-91

Sequence: 1 SEYNLLAEF 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 74574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	9	21	AA07882
2	40	95.2	19	22	AA07882
3	39	92.9	14	21	AA07888
4	38	90.5	9	21	AA07887
5	37	88.1	9	21	AA07879
6	36	85.7	9	21	AA07883
7	35	83.3	649	22	AB04111
8	34	81.0	9	19	AA082081
9	34	81.0	9	21	AA07874
10	34	81.0	9	21	AA07881
11	34	81.0	9	21	AA07894

12	34	81.0	9	22	AA073297	Protease indicator
13	34	81.0	10	18	AA08362	Beta-secretase sub
14	34	81.0	10	20	AA073756	Synthetic oligopep
15	34	81.0	10	21	AA069707	Beta-APP alpha-sec
16	34	81.0	10	22	AA010653	Human APP-Sw beta-
17	34	81.0	10	22	AA06898	Human APP-Sw beta-
18	34	81.0	10	22	AA06827	Human APP-Sw beta-
19	34	81.0	10	22	AA07226	Synthetic Asp2 rec
20	34	81.0	10	22	AA07226	Human beta-amyloid
21	34	81.0	10	22	AA07265	Human APP-Sw beta-
22	34	81.0	10	22	AA066575	Peptide 5-5'/SW, fo
23	34	81.0	10	22	AA061337	Synthetic peptide
24	34	81.0	11	22	AA075142	Swedish mutation p
25	34	81.0	11	22	AA075142	APP Swedish varian
26	34	81.0	11	22	AA07469	Asp 1 substrate se
27	34	81.0	16	21	AA06316	Asp2 substrate Swe
28	34	81.0	18	22	AA00609	Human beta-amyloid
29	34	81.0	19	22	AA00611	Beta-amyloid precu
30	34	81.0	20	19	AA082211	Amyloid precursor
31	34	81.0	20	21	AA06714	Fluorogenic protea
32	34	81.0	20	22	AA073229	Beta-APP alpha-sec
33	34	81.0	21	18	AA08361	Protease binding s
34	34	81.0	21	19	AA082186	Beta-secretase sub
35	34	81.0	21	20	AA033755	Synthetic oligopep
36	34	81.0	21	22	AA073203	Protease binding s
37	34	81.0	21	22	AA073204	Peptide 5-16'/SW, f
38	34	81.0	21	22	AA047265	Asp 1 substrate se
39	34	81.0	23	22	AA075148	Beta-secretase Swe
40	34	81.0	23	22	AA07474	Beta-secretase sub
41	34	81.0	30	18	AA08360	Synthetic oligopep
42	34	81.0	30	20	AA03754	Substrate for beta
43	34	81.0	30	21	AA07895	Peptide 26-4'/SW, f
44	34	81.0	30	22	AA047264	Mouse amyloid prec
45	34	81.0	32	17	AA04402	

#### ALIGNMENTS

RESULT 1	
AA07882	standard; peptide; 9 AA.
AA07882;	
14-NOV-2000	(first entry)
DE	A peptide fragment derived from beta-amyloid precursor protein.
KW	Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
KW	amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
KW	inhibitor.
XX	
OS	Homo sapiens.
XX	
PN	W0200047618-A2.
XX	
PD	17-AUG-2000.
XX	
PF	10-FEB-2000; 2000WO-US03819.
XX	
PR	10-FEB-1999; 99US-0119571.
PR	15-JUN-1999; 99US-0139172.
PA	(ELAN-) ELAN PHARM INC.
XX	
PI	Anderson JP, Bast G, Doane MT, Frigon N, John V, Power M;
XX	Stuha S, Tetsuno G, Tung J, Wang S, McConlogue L;
DR	WPI; 2000-533011/48.
XX	
PT	Purified beta-secretase protein used in assays to discover inhibitors
	which can be used for the treatment of amyloidogenic diseases e.g. "

PT Alzheimer's disease -  
 XX  
 PS Disclosure; Page 12; 121pp; English.  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 CC  
 SQ Sequence 9 AA:  
 Query Match 100.0%; Score 42; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SEVNLLAEF 9  
 DB 1 SEVNLLAEF 9  
 RESULT 2  
 AAE00612  
 ID AAE00612 standard; peptide; 19 AA.  
 XX  
 AC AAE00612;  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE pCBCL APP beta-secretase cleavage site (D/I) mutant.  
 XX  
 KW Amyloid precursor protein; APP; caspase; beta-secretase; cleavage site;  
 KW cysteine protease; apoptosis; caspase expression cassette; metastasis;  
 KW tumour; cathepsin B; urokinase; proliferation; gene therapy;  
 KW interdomain linker; Alzheimer's disease; mutant; mutain.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 12  
 FT /note= "Wild-type Asp substituted by Ile"  
 XX  
 PN WO200129232-A2.  
 XX  
 PD 26-APR-2001.  
 XX  
 PF 19-OCT-2000; 2000WO-US28941.  
 XX  
 PR 20-OCT-1999; 99US-0160559.  
 PR 14-AUG-2000; 2000US-0225564.  
 XX  
 PA (SCIO-) SCIOS INC.  
 XX  
 PI Cordell B, Li Y;  
 XX  
 DR WPI; 2001-290920/30.  
 XX  
 DR Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 PT proteolytic cleavage -  
 XX  
 PS Disclosure; Fig 29A; 116pp; English.  
 XX  
 CC The present amino acid sequence is a pCBCL plasmid amyloid precursor  
 CC protein (APP) beta-secretase cleavage site (AAE00611) (D/I) mutant. APP  
 CC sequence containing this mutation is less susceptible to beta-secretase

CC cleavage. APP beta-secretase cleavage site is used to construct an  
 CC artificially engineered chimeric cassette comprising human caspase-3 with  
 CC interdomain linker replaced by swedish mutant beta-secretase cleavage  
 CC site. This modified caspase-3 plays a pivotal role in Alzheimer's  
 CC disease. Caspases are a family of cysteine proteases, that participate  
 CC in the initiation and execution of apoptosis.  
 CC The present invention relates to a method for functional cloning of genes  
 CC encoding proteins or enzymes involved in proteolytic cleavage. The  
 CC invention is based on the use of caspase expression cassettes comprising  
 CC the coding sequence of a proteolytic cleavage site flanked by sequences  
 CC encoding two caspase subunits, separated by a cleavage site not associated  
 CC and a second caspase subunit, separated by a cleavage site not associated  
 CC in nature, is useful for cloning gene encoding enzymes involved in  
 CC proteolytic cleavage. An expression cassette containing fusion  
 CC polypeptide is used to identify a mutant cell line deficient in an  
 CC enzyme of interest and is also useful for diagnosis and suppression of  
 CC proliferation or metastases of a tumour cell characterised by  
 CC overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.  
 CC  
 SQ Sequence 19 AA:  
 Query Match 95.2%; Score 40; DB 22; Length 19;  
 Best Local Similarity 88.9%; Pred. No. 0.12;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SEVNLLAEF 9  
 DB 7 SEVNLLAEF 15  
 RESULT 3  
 AAB07888  
 ID AAB07888 standard; peptide; 14 AA.  
 XX  
 AC AAB07888;  
 XX  
 DT 14-NOV-2000 (first entry)  
 XX  
 DE A peptide fragment derived from beta-amyloid precursor protein.  
 XX  
 KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200047618-A2.  
 XX  
 PD 17-AUG-2000.  
 XX  
 PF 10-FEB-2000; 2000WO-US03819.  
 XX  
 PR 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX  
 PA (ELAN-) ELAN PHARM INC.  
 XX  
 PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX  
 DR WPI; 2000-533011/48.  
 XX  
 DR Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX  
 PS Disclosure; Page 12; 121pp; English.  
 XX  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque

CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC precursor sequence represents a peptide derived from beta-amyloid  
XX precursor protein

XX Sequence 14 AA;

Query Match 92.9%; Score 39; DB 21; Length 14;  
Best Local Similarity 88.9%; Pred. No. 0.14;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
| | | | |  
Db 6 SEVNLVAEF 14

## RESULT 4

AAB07887  
ID AAB07887 standard; peptide; 9 AA.

XX AAB07887;

DT 14-NOV-2000 (first entry)

DE A peptide fragment derived from beta-amyloid precursor protein.

KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
XX inhibitor.

OS Homo sapiens.

PN WO200047618-A2.

PD 17-AUG-2000.

PF 10-FEB-2000; 2000WO-US03819.

PR 10-FEB-1999; 99US-0119571.

PR 15-JUN-1999; 99US-0139172.

PA (ELAN-) ELAN PHARM INC.

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

DR WPI: 2000-533011/48.

XX Purified beta-secretase protein used in assays to discover inhibitors  
XX which can be used for the treatment of amyloidogenic diseases e.g.  
XX Alzheimer's disease.

PS Disclosure; Page 12; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
XX beta-amyloid precursor protein to produce beta-amyloid peptide. This  
XX enzyme is therefore implicated in the production of amyloid plaque  
XX components which accumulate in the brains of individuals afflicted with  
XX Alzheimer's disease. Inhibitors of beta-secretase are administered to  
XX a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
XX disease-like pathology to test if they maintain or improve cognitive  
XX ability or reduce the plaque burden. The compounds are used for the  
XX treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
XX present sequence represents a peptide derived from beta-amyloid  
XX precursor protein

XX Sequence 9 AA;

Query Match 90.5%; Score 38; DB 21; Length 9;

Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
| | | | |  
Db 1 SEVNLVAEF 9

## RESULT 5

AAB07879  
ID AAB07879 standard; peptide; 9 AA.

XX AAB07879;

DT 14-NOV-2000 (first entry)

DE A peptide fragment derived from beta-amyloid precursor protein.

KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
XX inhibitor.

OS Homo sapiens.

PN WO200047618-A2.

PD 17-AUG-2000.

PF 10-FEB-2000; 2000WO-US03819.

PR 10-FEB-1999; 99US-0119571.

PR 15-JUN-1999; 99US-0139172.

PA (ELAN-) ELAN PHARM INC.

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

DR WPI: 2000-533011/48.

XX Purified beta-secretase protein used in assays to discover inhibitors  
XX which can be used for the treatment of amyloidogenic diseases e.g.  
XX Alzheimer's disease.

PS Disclosure; Page 12; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
XX beta-amyloid precursor protein to produce beta-amyloid peptide. This  
XX enzyme is therefore implicated in the production of amyloid plaque  
XX components which accumulate in the brains of individuals afflicted with  
XX Alzheimer's disease. Inhibitors of beta-secretase are administered to  
XX a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
XX disease-like pathology to test if they maintain or improve cognitive  
XX ability or reduce the plaque burden. The compounds are used for the  
XX treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
XX present sequence represents a peptide derived from beta-amyloid  
XX precursor protein

XX Sequence 9 AA;

Query Match 88.1%; Score 37; DB 21; Length 9;  
Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
| | | | |  
Db 1 SEVNLVAEF 9

## RESULT 6

AAB07883  
ID AAB07883 standard; peptide; 9 AA.

XX

AC	AAB07883;
XX	14-NOV-2000 (first entry)
DT	
XX	A peptide fragment derived from beta-amyloid precursor protein.
DE	
KW	Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
KW	amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
KW	inhibitor.
XX	
OS	Homo sapiens.
XX	
PN	MO200047618-A2.
XX	
PD	17-AUG-2000.
XX	
PF	10-FEB-2000; 2000WO-US03819.
PR	
XX	10-FEB-1999; 99US-0119571.
PR	15-JUN-1999; 99US-0139172.
XX	
PA	(ELAN-) ELAN PHARM INC.
PX	
PI	Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M:
PI	Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;
XX	
DR	WPI: 2000-533011/48.
XX	
PT	Purified beta-secretase protein used in assays to discover inhibitors
PT	which can be used for the treatment of amyloidogenic diseases e.g.
PT	Alzheimer's disease -
XX	
XX	Disclosure: Page 12; 121pp; English.
XX	
CC	The specification describes a beta-secretase enzyme. The enzyme cleaves
CC	beta-amyloid precursor protein to produce beta-amyloid peptide. This
CC	enzyme is therefore implicated in the production of amyloid plaque
CC	components which accumulate in the brains of individuals afflicted with
CC	Alzheimer's disease. Inhibitors of beta-secretase are administered to
CC	a mammalian subject e.g. with Alzheimer's disease or Alzheimer's
CC	disease-like pathology to test if they maintain or improve cognitive
CC	ability or reduce the plaque burden. The compounds are used for the
CC	treatment of amyloidogenic diseases e.g. Alzheimer's disease. The
CC	present sequence represents a peptide derived from beta-amyloid
CC	precursor protein
XX	
SQ	Sequence 9 AA;
XX	
QY	Query Match 85.7%; Score 36; DB 21; Length 9;
DB	Best Local Similarity 88.9%; Pred. No. 6,4e+05; Indels 0; Gaps 0;
	Matches 8; Conservative 0; Mismatches 1;
OY	1 SEVNLLAEF 9
ID	1 SEVKLLAEF 9
XX	
RESULT 7	
ID	ABB64111 standard; Protein: 649 AA.
XX	
AC	ABB64111;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 19125.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.

PD	XX	27-SEP-2001.
PD	XX	
PF	XX	23-MAR-2001; 2001WO-US09231.
PR	XX	23-MAR-2000; 2000US-191637P.
PR	XX	11-JUL-2000; 2000US-0614150.
PA	XX	(PEKE ) PE CORP NY.
PI	XX	Venter JC, Adams M, Li PWD, Myers EM;
DR	XX	WPI: 2001-656860/75.
DR	XX	N-PSDB; ABL08214.
PT	XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	XX	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	XX	interactions -
PS	XX	Disclosure: SEQ ID NO 19125; 21pp + Sequence Listing; English.
PS	XX	
CC	XX	The invention relates to an isolated nucleic acid detection reagent
CC	XX	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	XX	useful in developmental biology and in elucidating cell signalling and
CC	XX	cell-cell interactions in higher eukaryotes for the development of
CC	XX	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	XX	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	XX	sequences (ABBS7737-ABBS7072).
CC	XX	(ABBS7737-ABBS7072).
CC	XX	The sequence data for this patent did not form part of the printed
CC	XX	specification, but was obtained in electronic format directly from WIPO
CC	XX	at ftp.wipo.int/pub/published_pct_sequences.
SO	XX	Sequence 649 AA:
QY	1 SEVNILAE 8	
	:	
Db	131 SEINILAE 138	
RESULT 8		
AAW82081		
ID	AAW82081 standard; peptide: 9 AA.	
XX		
AC	AAW82081:	
XX		
DT	18-FEB-1999 (first entry)	
DE	Fluorogenic protease indicator protease binding peptide #59.	
KW	Protease activity; fluorophore; detection; fluorogenic; cellular uptake;	
KW	conformation change.	
XX		
OS	Synthetic.	
XX		
PN	WO9837226-A1.	
XX		
PD	27-AUG-1998.	
XX		
PF	20-FEB-1998; 98WO-US03000.	
XX		
PR	20-FEB-1997; 97US-0802981.	
XX		
PA	(ONCO-) ONCOIMMUNIN INC.	
XX		
PI	Komoriya A, Packard BS;	
XX		
DR	WPI; 1998-467579/40.	
XX		

PT New fluorogenic compositions - containing 2 fluorophores separated  
PT by a peptide comprising a protease binding site, used for detecting  
XX protease activity in samples.  
XX  
PS Claim 4; Page 77; 90pp; English.  
XX  
CC AAM82023-W82240 are peptides used in the construction of a fluorogenic  
CC composition which is used for the detection of protease activity in  
CC biological samples. The products can be used for the detection of  
CC conformation changes in nucleic acids, oligosaccharides,  
CC polysaccharides, steroids or polymers. In addition, attachment of a  
CC glycoproteins, lipids, phospholipids, glycolipids,  
CC hydrophobic group to a molecule can be used to enhance uptake by cells.  
CC The composition is composed of P - peptide comprising a protease binding  
CC site for the protease, F1, F2 peptides - fluorophores where F1 is  
CC attached to the amino terminal amino acid and F2 is attached to the  
CC carboxyl terminal amino acid and S1, S2 peptides - when present, are  
CC peptide spacers where S1, when present, is attached to the amino terminal  
CC acid, and S2, when present, is attached to the carboxyl terminal amino  
CC acid.  
XX  
SQ Sequence 9 AA:  
XX  
Query Match 81.0%; Score 34; DB 19; Length 9;  
Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 SEVNLLAEF 9  
IIII III  
DB 1 SEVNLLDAEF 9  
XX  
RESULT 9  
AAB07874  
ID AAB07874 standard; peptide; 9 AA.  
XX  
AC AAB07874;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE A peptide fragment derived from beta-amyloid precursor protein.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
XX  
PN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI; 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
XX Alzheimer's disease -  
XX  
PS Disclosure; Page 12; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:  
XX  
Query Match 81.0%; Score 34; DB 19; Length 9;  
Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:  
XX  
Query Match 81.0%; Score 34; DB 21; Length 9;  
Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 SEVNLLAEF 9  
IIII III  
DB 1 SEVNLLDAEF 9  
XX  
RESULT 10  
AAB07881  
ID AAB07881 standard; peptide; 9 AA.  
XX  
AC AAB07881;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE A peptide fragment derived from beta-amyloid precursor protein.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
XX  
PN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI; 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
XX Alzheimer's disease -  
XX  
PS Disclosure; Page 12; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:  
XX  
Query Match 81.0%; Score 34; DB 21; Length 9;

Best Local Similarity 77.8%; Pred. No. 6.4e+05;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
| | | : | | | |  
Db 1 SEVNLAEF 9

RESULT 11

AAB07894 standard; Peptide: 9 AA.

AAB07894;

14-NOV-2000 (first entry)

Substrate for beta-secretase enzyme.

Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
inhibitor.

Synthetic.

Key Location/Qualifiers  
Cleavage-site 5..6

WO200047618-A2.

17-AUG-2000.

10-FEB-2000; 2000WO-US03819.

10-FEB-1999; 99US-0119571.  
15-JUN-1999; 99US-0139172.

(ELAN-) ELAN PHARM INC.

Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;  
Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

WPI: 2000-533011/48.

Purified beta-secretase protein used in assays to discover inhibitors  
which can be used for the treatment of amyloidogenic diseases e.g.

Alzheimer's disease -

Example 4; Page 71; 121pp; English.

The specification describes a beta-secretase enzyme. The enzyme cleaves  
beta-amyloid precursor protein to produce beta-amyloid peptide. This  
enzyme is therefore implicated in the production of amyloid plaque  
components which accumulate in the brains of individuals afflicted with  
Alzheimer's disease. Inhibitors of beta-secretase are administered to  
a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
disease-like pathology to test if they maintain or improve cognitive  
ability or reduce the plaque burden. The compounds are used for the  
treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
present sequence represents a peptide substrate used to test the  
activity of beta-secretase enzyme.

Sequence 9 AA:

Query Match 81.0%; Score 34; DB 21; Length 9;

Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
| | | | | | | | | |

Db 1 SEVNLAEF 9

RESULT 12

AAG73297  
ID AAG73297 standard; Peptide: 9 AA.

AAG73297;

14-AUG-2001 (first entry)

Protease indicator compound peptide #26.

Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
viral infection; cancer metastasis; emphysema; arthritis;  
thrombosis; haemophilia.

Synthetic.

WO200118238-A1.

15-MAR-2001.

11-SEP-2000; 2000WO-US24882.

10-SEP-1999; 99US-0394019.

(ONCO-) ONCOIMMUNIN INC.

Komoriya A, Packard BS;

WPI: 2001-389573/41.

New fluorogenic compositions whose fluorescence level increases in the  
presence of active proteases, useful for detecting and localizing  
protease activity in biological samples, particularly in frozen tissue  
samples -

Claim 1; Page 70; 86pp; English.

The present invention describes fluorogenic compositions which can be  
used for the detection of protease activity. This can be useful as an  
indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
thrombosis and arthritis. The fluorogenic compositions comprise a  
peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
peptide is cleaved by a protease and the fluorophores can then be  
detected. The present sequence is one of the peptides described in the  
exemplification of the invention.

Sequence 9 AA:

Query Match 81.0%; Score 34; DB 22; Length 9;

Best Local Similarity 88.9%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLAEF 9  
| | | | | | | | | |

Db 1 SEVNLAEF 9

RESULT 13

AAW08362  
ID AAW08362 standard; Peptide: 10 AA.

AAW08362;

05-SEP-1997 (first entry)

Beta-secretase substrate #3.

Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;  
alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.

Synthetic.

Key Location/Qualifiers

Misc-difference 1

FT /note- "acetylated"  
XX  
PN W09640885-A2.  
XX  
PD 19-DEC-1996.  
XX  
PE 07-JUN-1996; 96WO-US09985.  
XX  
PR 07-JUN-1995; 95US-0485152.  
PR 07-JUN-1995; 95US-0480498.  
XX  
PA (ATHE-) ATHENA NEUROSCIENCES INC.  
XX  
PI Anderson JP, Chrysler SMS, Jacobson-croak KL, Keim PS;  
PI McConlogue LC, Sinha S, Tan H;  
DR WPI; 1997-052304/05.  
XX  
PT Beta-secretase which specifically cleaves beta-amyloid precursor  
PT protein - useful to screen for inhibitors useful in treatment of  
PT Alzheimer's disease  
XX  
PS Disclosure; Page 45; 92pp; English.  
XX  
CC AAW08359-W08362 represent substrates for the enzyme of the invention.  
CC The enzyme of the invention is beta-secretase, and specifically cleaves  
CC beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP  
CC is thought to occur via cleavage between residues 16 and 17 of the  
CC beta-amyloid peptide region by an alpha-secretase. Pathogenic processing  
CC is thought to occur by beta-secretase cleavage of beta-APP.  
CC Beta-secretase activity can be detected and measured using a method of  
CC the invention, which detects at least one of the beta-secretase cleavage  
CC products formed on cleavage. The method can be used to determine whether  
CC a test substance inhibits proteolytic cleavage, by beta-secretase, of  
CC beta-APP. Compounds effective to at least partially inhibit  
CC beta-secretase activity can be used to inhibit cleavage of beta-APP in  
CC cells or mammalian hosts. Isolation and purification of beta-secretase  
CC will permit chemical modelling of a critical event in the pathology of  
CC Alzheimer's disease.  
SQ Sequence 10 AA;  
XX  
Query Match 81.0%; Score 34; DB 18; Length 10;  
Best Local Similarity 88.9%; Pred. No. 1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SEVNLLAEF 9  
DB 1 SEVNLLAEF 9  
XX  
RESULT 14  
AAV33756  
ID AAV33756 standard; Protein; 10 AA.  
XX  
AC AAV33756;  
XX  
DE 09-NOV-1999 (first entry)  
XX  
DE Synthetic oligopeptide 5-5'SW.  
XX  
KW Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;  
KW Alzheimer's disease; measure activity; cleavage site.  
XX  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FT Modified-site 1 /note- "N-terminal Ser is acetylated"  
XX  
XX US5942400-A.  
XX  
XX 24-AUG-1999.

XX  
PE 07-JUN-1996; 96US-0659984.  
XX  
PR 07-JUN-1996; 96US-0659984.  
PR 07-JUN-1995; 95US-0480498.  
PR 07-JUN-1995; 95US-0485152.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Jacobson-Croak KL, Sinha S;  
PI WPI; 1999-517417/43.  
DR  
XX  
PT A method for detecting human beta-secretase cleavage of polypeptides  
PT useful for identifying beta-secretase inhibitors  
XX  
PS Examples; Column 30; 43pp; English.  
XX  
CC Sequences AAV33752-Y33756 are synthetic oligopeptides used for measuring  
CC the activity of beta-secretase (AAV33741). Beta-secretase is capable of  
CC cleaving beta-amyloid protein precursor (APP) (AAV33742). These  
CC synthetic peptides contain the cleavage site of APP. Beta-secretase and  
CC APP are used in a method for detecting human beta-secretase cleavage of  
CC polypeptides and for identifying beta-secretase inhibitors. Inhibition  
CC of beta-secretase activity would be useful for chemical modelling of a  
CC critical event in the pathology of Alzheimer's disease. Inhibitors of  
CC beta-secretase would be useful for the prevention and treatment of  
CC Alzheimer's disease and Down's Syndrome.  
SQ Sequence 10 AA;  
XX  
Query Match 81.0%; Score 34; DB 20; Length 10;  
Best Local Similarity 88.9%; Pred. No. 1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SEVNLLAEF 9  
DB 1 SEVNLLAEF 9  
XX  
RESULT 15  
AAV69707  
ID AAV69707 standard; peptide; 10 AA.  
XX  
AC AAV69707;  
XX  
DE 11-APR-2000 (first entry)  
XX  
DE Beta-APP alpha-secretase substrate [NED]-APP(-5,+5).  
XX  
KW Neotropic; neuroprotective; beta-amyloid precursor protein; metabolism;  
KW cleavage site; beta-secretase; neurodegenerative disease;  
KW Alzheimer's disease.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN W09964587-A1.  
XX  
PD 16-DEC-1999.  
XX  
PE 04-JUN-1999; 99WO-FR01326.  
XX  
PR 05-JUN-1998; 98FR-0007068.  
PR 31-MAR-1999; 99US-0122599.  
XX  
PA (RHON) RHONE-POULENC RORER SA.  
PA (UYPA-) UNIV CURIE PARIS VI P & M.  
XX  
PI Rholam M, Munoz-Gimenez N, Moutaouakil M, Cohen P, Bertrand P;  
PI WPI; 2000-097537/08.  
DR  
XX

PT Polypeptide with beta-secretase activity, specific for wild-type  
amyloid precursor protein, useful in treating Alzheimer's disease

XX  
PS Example 3; Page 24; 44pp; French.

XX Peptides AAV69702-Y69718 represent synthetic peptide substrates for a  
CC novel polypeptide with beta-secretase activity that can cleave  
CC specifically the natural beta-amyloid precursor protein (bAPP). Normal  
CC cleavage of the protein occurs between amino acids Met596-Asp597 and  
CC Val636-Ile637 (positions 4-5 and 44-45 of AAV69701). The novel  
CC polypeptide is used to identify agents that interact specifically with  
CC it. These agents regulate metabolism of APP, particularly they slow down  
CC or reduce production of beta-amyloid, so can be used to treat  
CC neurodegenerative diseases, particularly Alzheimer's disease.

XX  
SO Sequence 10 AA;

Query Match 81.0%; Score 34; DB 21; Length 10;  
Best Local Similarity 88.9%; Pred. No. 1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SEVNLLAEF 9  
|||||  
Db 1 SEVNLDAEF 9

Search completed: October 30, 2002, 12:27:09  
Job time : 3.60442 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2002, 12:24:20 / Search time 51.4447 Seconds  
(without alignments)  
171.400 Million cell updates/sec

Title: US-09-724-571-75

Sequence: 1 MVDLRKSGSGGYVEMTVG.....VFEDRRKRIGFAVSACHVH 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

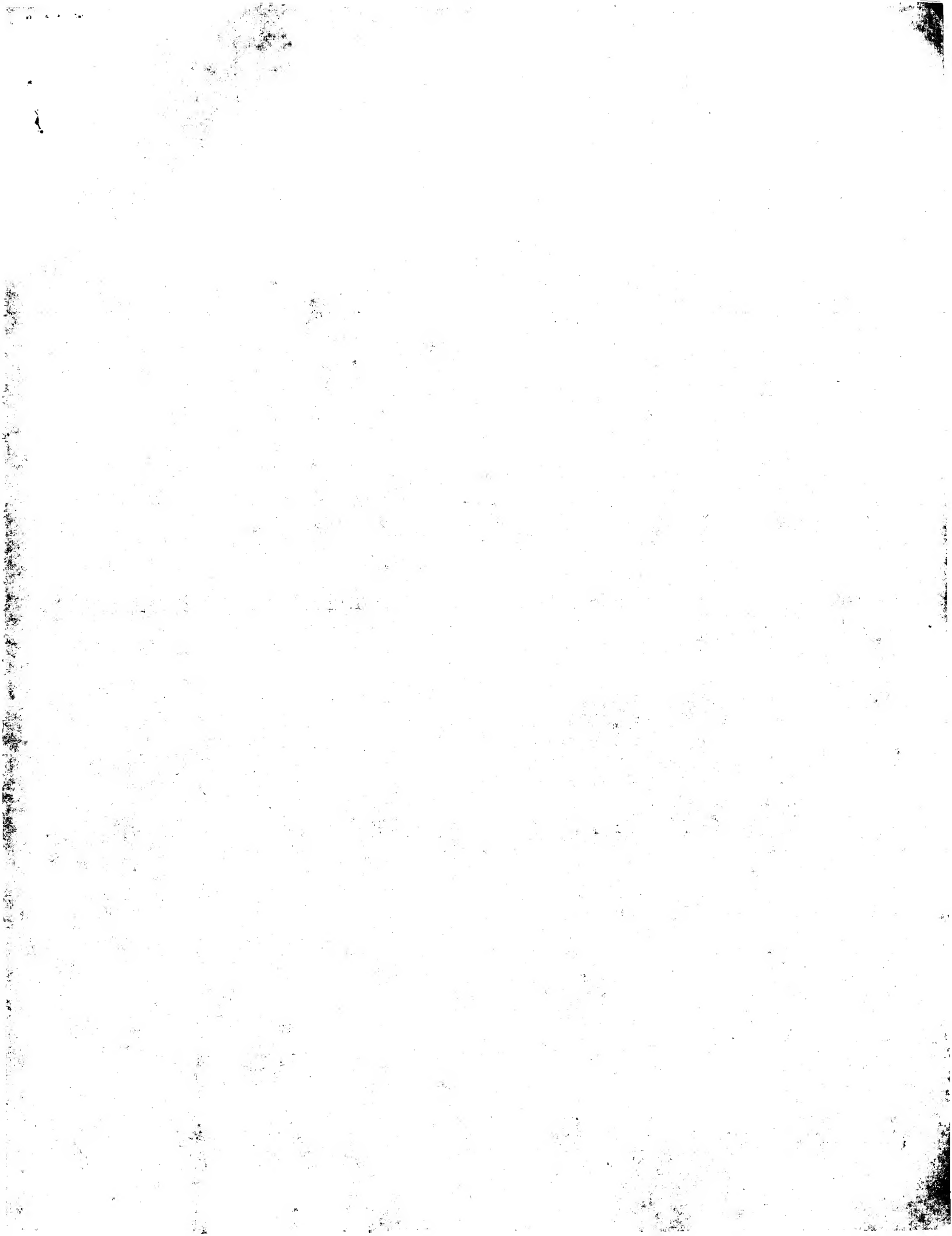
## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1901	99.7	501	4	US-09-009-191-2 Sequence 2, Appl1
2	1890	99.1	774	4	US-09-009-191-4 Sequence 4, Appl1
3	1106	58.0	518	3	US-08-999-723-2 Sequence 2, Appl1
4	1106	58.0	518	4	US-09-434-427-2 Sequence 2, Appl1
5	1088	57.1	514	4	US-09-912-484-2 Sequence 2, Appl1
6	1088	57.1	514	4	US-08-912-484-2 Sequence 2, Appl1
7	298.5	15.7	396	4	US-08-208-007A-13 Sequence 13, Appl1
8	298.5	15.7	396	4	US-08-032-523-9 Sequence 9, Appl1
9	295.5	15.5	412	1	US-08-208-007A-12 Sequence 12, Appl1
10	295.5	15.5	412	4	US-08-974-691-4 Sequence 4, Appl1
11	279.5	14.7	458	6	Patent No. 5217891 Sequence 6, Appl1
12	273.5	14.3	409	1	US-09-640-305-6 Sequence 6, Appl1
13	273.5	14.3	409	1	US-08-360-673-6 Sequence 8, Appl1
14	273.5	14.3	427	2	US-08-846-021A-8 Sequence 2, Appl1
15	270	14.2	410	1	US-08-088-633-2 Sequence 2, Appl1
16	270	14.2	410	1	US-08-088-633-2 Sequence 2, Appl1
17	270	14.2	410	1	US-08-245-756-2 Sequence 2, Appl1
18	270	14.2	410	1	US-08-441-750-2 Sequence 2, Appl1
19	270	14.2	410	5	PCR-US92-02521-2 Sequence 2, Appl1
20	249	13.1	349	4	US-09-032-523-3 Sequence 2, Appl1
21	227	11.9	398	1	US-08-328-314-2 Sequence 2, Appl1
22	227	11.9	398	1	US-08-328-314-2 Sequence 2, Appl1
23	212	11.1	419	4	US-08-974-691-3 Sequence 3, Appl1
24	210	11.0	397	3	US-09-079-415-2 Sequence 2, Appl1
25	195.5	10.3	427	1	US-07-958-222A-2 Sequence 2, Appl1
26	194.5	10.2	430	1	US-08-535-237-2 Sequence 2, Appl1
27	193	10.1	330	3	US-08-113-735-1 Sequence 1, Appl1

28	193	10.1	419	3	US-08-115-753-2 Sequence 2, Appl1
29	193	10.1	419	3	US-08-115-753-3 Sequence 33, Appl1
30	184	9.6	445	4	US-08-974-691-6 Sequence 6, Appl1
31	184	9.6	451	4	US-08-974-691-2 Sequence 2, Appl1
32	180	9.4	420	4	US-08-008-271A-4 Sequence 4, Appl1
33	180	9.4	420	4	US-08-974-691-8 Sequence 8, Appl1
34	177.5	9.3	395	1	US-08-723-938-3 Sequence 3, Appl1
35	177.5	9.3	395	2	US-09-080-538-3 Sequence 3, Appl1
36	149	7.8	437	4	US-09-353-332-2 Sequence 2, Appl1
37	129.5	6.8	140	3	US-09-211-631-13 Sequence 13, Appl1
38	129.5	6.8	140	4	US-09-265-628-13 Sequence 13, Appl1
39	129.5	6.8	140	4	US-09-001-141-11 Sequence 11, Appl1
40	129.5	6.8	140	4	US-09-532-803-6 Sequence 6, Appl1
41	129.5	6.8	140	4	US-09-653-403-14 Sequence 14, Appl1
42	97	5.1	1030	4	US-09-091-117-2 Sequence 2, Appl1
43	95.5	5.0	280	4	US-09-160-246-14 Sequence 14, Appl1
44	85	4.5	377	2	US-08-853-659A-41 Sequence 41, Appl1
45	84.5	4.4	1097	2	US-08-680-326-39 Sequence 39, Appl1

## ALIGNMENTS

RESULT 1  
US-09-009-191-2  
Sequence 2, Application US/09009191  
Patent No. 6319689  
GENERAL INFORMATION:  
APPLICANT: POWELL, DAVID  
APPLICANT: CHAPMAN, CONRAD  
APPLICANT: MORPHY, KAY  
APPLICANT: SMITH, TRUDI  
TITLE OF INVENTION: ASP2  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,191  
FILING DATE: 20-JAN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: UK 9701684.4  
FILING DATE: 28-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70368  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-009-191-2  
Query Match 99.7%: Score 1901; DB 4; Length 501;  
Best Local Similarity 99.7%: Pred. No. 1.4e-187;  
Matches 360; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



```

QY      1 MYDNLRGKSGGCIYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFLHRYRQROLSSTYR 60
Db      63 MYDNLRGKSGGCIYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFLHRYRQROLSSTYR 122
QY      61 DLKRGVYPTGKKEGELGDLVSIPIHGPVNTVYRANIAATESDKFFINSNMEGILGL 120
Db      123 DLKRGVYPTGKKEGELGDLVSIPIHGPVNTVYRANIAATESDKFFINSNMEGILGL 182
QY      121 AYAEIARPDSDLPEPFDLSVKTQHPNLFSLQCGAGFPLNQSIVLASVSGSMITIGIDH 180
Db      183 AYAEIARPDSDLPEPFDLSVKTQHPNLFSLQCGAGFPLNQSIVLASVSGSMITIGIDH 242
QY      181 SLTGSIMTPTPIRRRWYEVILVREINQDGLKMDCKEYNDKSIYDSGTTNLRPKV 240
Db      243 SLTGSIMTPTPIRRRWYEVILVREINQDGLKMDCKEYNDKSIYDSGTTNLRPKV 302
QY      241 EAAVKSIAASTEKFPDGFVLGEOVLCWQAGTTPMNIFFVLSLYLGEVYTNOSFRITIL 300
Db      303 EAAVKSIAASTEKFPDGFVLGEOVLCWQAGTTPMNIFFVLSLYLGEVYTNOSFRITIL 362
QY      301 POQYLRPEVDVATSDDCYKFAISQSTGYVMGAVIMEGFYVDFDRAKRIGFAVSACHV 360
Db      363 POQYLRPEVDVATSDDCYKFAISQSTGYVMGAVIMEGFYVDFDRAKRIGFAVSACHV 422
QY      361 H 361
Db      423 H 423

```

```

RESULT 2
US-09-009-191-4
; Sequence 4, Application US/09009191
; Patent No. 6319689

```

```

; GENERAL INFORMATION:
; APPLICANT: POWELL, DAVID
; APPLICANT: CHAPMAN, CONRAD
; APPLICANT: MORPHY, KAY
; APPLICANT: SMITH, TRUDI
; TITLE OF INVENTION: ASP2
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,191
; FILING DATE: 20-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9701684.4
; FILING DATE: 28-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70368
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELERX: 846169

```

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: amino acid
; COMPLETENESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-009-191-4

```

```

Query Match          99.1%; Score 1890; DB 4; Length 774;
Best Local Similarity 99.2%; Pred. No. 3,76-186;
Matches 358; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 MYDNLRGKSGGCIYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFLHRYRQROLSSTYR 60
Db      6 MYDNLRGKSGGCIYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFLHRYRQROLSSTYR 65
QY      61 DLKRGVYPTGKKEGELGDLVSIPIHGPVNTVYRANIAATESDKFFINSNMEGILGL 120
Db      66 DLKRGVYPTGKKEGELGDLVSIPIHGPVNTVYRANIAATESDKFFINSNMEGILGL 125
QY      121 AYAEIARPDSDLPEPFDLSVKTQHPNLFSLQCGAGFPLNQSIVLASVSGSMITIGIDH 180
Db      126 AYAEIARPDSDLPEPFDLSVKTQHPNLFSLQCGAGFPLNQSIVLASVSGSMITIGIDH 185
QY      181 SLTGSIMTPTPIRRRWYEVILVREINQDGLKMDCKEYNDKSIYDSGTTNLRPKV 240
Db      186 SLTGSIMTPTPIRRRWYEVILVREINQDGLKMDCKEYNDKSIYDSGTTNLRPKV 245
QY      241 EAAVKSIAASTEKFPDGFVLGEOVLCWQAGTTPMNIFFVLSLYLGEVYTNOSFRITIL 300
Db      246 EAAVKSIAASTEKFPDGFVLGEOVLCWQAGTTPMNIFFVLSLYLGEVYTNOSFRITIL 305
QY      301 POQYLRPEVDVATSDDCYKFAISQSTGYVMGAVIMEGFYVDFDRAKRIGFAVSACHV 360
Db      306 POQYLRPEVDVATSDDCYKFAISQSTGYVMGAVIMEGFYVDFDRAKRIGFAVSACHV 365
QY      361 H 361
Db      366 H 366

```

```

RESULT 3
US-08-999-723-2
; Sequence 2, Application US/08999723A
; Patent No. 6025180

```

```

; GENERAL INFORMATION:
; APPLICANT: Powell, David J.
; APPLICANT: Soutchan, Christopher
; APPLICANT: Chapman, Conrad G.
; APPLICANT: Evans, Joanne R.
; TITLE OF INVENTION: Asp1
; FILE REFERENCE: GH70262
; CURRENT APPLICATION NUMBER: US/08/999,723A
; CURRENT FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-999-723-2

```

```

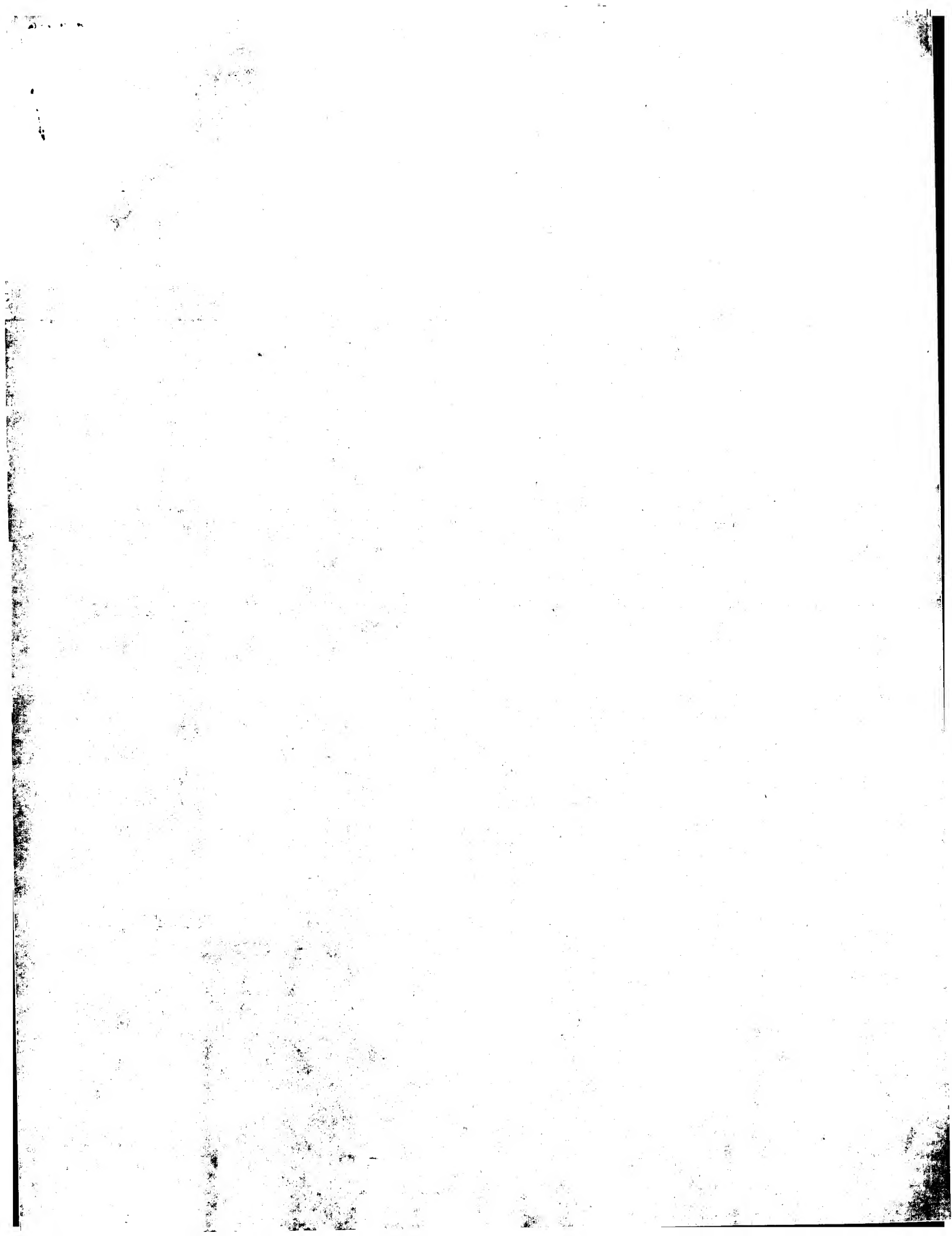
Query Match          58.0%; Score 1106; DB 3; Length 518;
Best Local Similarity 56.4%; Pred. No. 1,5e-105;
Matches 202; Conservative 60; Mismatches 92; Indels 4; Gaps 2;

```

```

QY      1 MYDNLRGKSGGCIYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFLHRYRQROLSSTYR 60
Db      80 MYDNLRGKSGGCIYVEMTVGSPPTLNILVDTGSSNFAVGAAPHFLHRYRQROLSSTYR 139
QY      61 DLKRGVYPTGKKEGELGDLVSIPIHGPVNTVYRANIAATESDKFFINSNMEGILGL 120
Db      140 SKGFDTVAKYITGSGWTGPGEDLVITPKGPNLSLVNIAATFESBNFELPKIKWNGILGL 199
QY      121 AYAEIARPDSDLPEPFDLSVKTQHPNLFSLQCGAGFPLNQSIVLASVSGSMITIGIDH 180
Db      200 AYATLAKPSSSLETFEFDLSVKTQHPNLFSLQCGAGFPLNQSIVLASVSGSMITIGIDH 256

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:16:14 ; Search time 56.55 seconds

(without alignments)  
799.417 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 1 ETDPEEPERGRGRSFVEMVD.....GPFVTLDMEDCGYNIPQIDE 407

Sequence: 1 ETDPEEPERGRGRSFVEMVD.....GPFVTLDMEDCGYNIPQIDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_032802:\*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2156	100.0	415	21	AA807899
2	2156	100.0	453	21	AA88438
3	2156	100.0	453	22	AAE10643
4	2156	100.0	453	22	AAE06872
5	2156	100.0	453	22	AAU06616
6	2156	100.0	453	22	AAU07215
7	2156	100.0	453	22	AAE02594
8	2156	100.0	456	21	AA807897
9	2156	100.0	459	21	AA88439
10	2156	100.0	459	22	AAE10643
11	2156	100.0	459	22	AAE06873

12	2156	100.0	459	22	AAU06617	Human-pro-Asp 2(a)
13	2156	100.0	459	22	AAU07216	Human aspartyl pro
14	2156	100.0	459	22	AAE02595	Human-asp-2(a) del
15	2156	100.0	460	21	AA807898	Amino acid sequenc
16	2156	100.0	488	22	AAE65572	Human memapsin 2.
17	2156	100.0	488	22	AAE61334	Human memapsin 2.
18	2156	100.0	501	21	AA94767	Memapsin 2 protein
19	2156	100.0	501	21	AA807896	Human beta-secreta
20	2156	100.0	503	22	AAE6573	Amino acid sequenc
21	2156	100.0	503	22	AAE61335	Human pro-memapsin
22	2151	99.8	433	21	AA88433	T7 promoter and ve
23	2151	99.8	433	22	AAE10640	Human-pro-asp-2(a)
24	2151	99.8	433	22	AAE06870	Human-pro-asp 2(a)
25	2151	99.8	433	22	AAE06614	Human-pro-asp2(a)
26	2151	99.8	433	22	AAU06614	Human-pro-asp 2(a)
27	2151	99.8	433	22	AAE02592	T7-human aspartyl
28	2151	99.8	446	21	AA88431	Human-pro-asp-2(a)
29	2151	99.8	446	22	AAE10638	Human-pro-asp-2(a)
30	2151	99.8	446	22	AAE06868	Human-pro-asp-2(a)
31	2151	99.8	446	22	AAU06612	Human-pro-asp-2(a)
32	2151	99.8	446	22	AAE02590	T7-caspase-human-p
33	2151	99.8	446	22	AAE02590	T7-caspase-human-p
34	2151	99.8	459	21	AA88432	T7-caspase-human-p
35	2151	99.8	459	22	AAE10639	T7-caspase-human-p
36	2151	99.8	459	22	AAE06869	T7-caspase-human-p
37	2151	99.8	459	22	AAU06613	T7-caspase-human-p
38	2151	99.8	459	22	AAU07212	T7-caspase-human-p
39	2151	99.8	459	22	AAE02591	T7-caspase-human-p
40	2151	99.8	501	21	AA88425	Human aspartyl pro
41	2151	99.8	501	21	AAE10629	Human aspartyl pro
42	2151	99.8	501	22	AAE06859	Human aspartyl pro
43	2151	99.8	501	22	AAE06603	Human aspartyl pro
44	2151	99.8	501	22	AAU07202	Human aspartyl pro
45	2151	99.8	501	22	AAE02581	Human aspartyl pro

## ALIGNMENTS

RESULT 1

AA807899 standard; Protein; 415 AA.

AC AA807899;

DT 14-NOV-2000 (first entry)

XX

XX Amino acid sequence of a human beta-secretase enzyme fragment.

XX

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;

KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

KW inhibitor.

XX

XX Homo sapiens.

OS

XX

PN W0200047618-A2.

PD 17-AUG-2000.

XX

XX 10-FEB-2000: 2000WO-US03819.

PF

XX 10-FEB-1999: 99US-0119571.

PR 15-JUN-1999: 99US-0139172.

XX

PA (ELAN-) ELAN PHARM INC.

XX

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX WPI: 2000-533011/48.

DR

XX Purified beta-secretase protein used in assays to discover inhibitors

PT which can be used for the treatment of amyloidogenic diseases e.g.



XX 10-DEC-2001 (first entry)  
 DT Human-Asp 2(a) protein lacking transmembrane domain.  
 DE  
 XX  
 KW Human: aspartyl protease 2a; Asp2a: amyloid precursor protein; APP;  
 KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;  
 KW Alzheimer plaque; neuronal loss; proteolytic; neurotrophic; neuroprotective.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 214 /note="Encoded by CAC"  
 FT  
 PN GB357767-A.  
 PD 04-JUL-2001.  
 XX  
 PF 22-SEP-2000; 2000GB-0023315.  
 XX  
 PR 23-SEP-1999; 99US-0155493.  
 PR 23-SEP-1999; 99US-0404133.  
 PR 23-SEP-1999; 99MO-US20881.  
 PR 13-OCT-1999; 99US-0416801.  
 PR 06-DEC-1999; 99US-0169232.  
 XX  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 PI  
 PI Bienkowski MJ, Gurney M;  
 DR WPI: 2001-444208/48.  
 DR N-PSDB; AAD17878.  
 XX  
 PT Polypeptide comprising fragments of human aspartyl protease with  
 PT amyloid precursor protein processing activity and alpha-secretase  
 PT activity, for identifying modulators useful in treating Alzheimer's  
 XX disease -  
 XX  
 PS Example 10; Fig 11; 187pp; English.  
 CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified  
 CC Asp1 proteins which lack transmembrane domain or amino terminal  
 CC domain or cytoplasmic domain and retains alpha-secretase activity  
 CC and amyloid protein precursor (APP) processing activity. The proteins  
 CC of the invention are useful for assaying hu-Asp1 alpha-secretase  
 CC activity, which in turn is useful for identifying modulators of  
 CC hu-Asp1 alpha-secretase activity, where modulators that increase  
 CC hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's  
 CC disease (AD) which causes progressive dementia with consequent  
 CC formation of amyloid plaques, neurofibrillary tangles, gliosis and  
 CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying  
 CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with  
 CC the substrate under acidic conditions and determining the level of  
 CC hu-Asp1 proteolytic activity. The present sequence is human Asp 2(a)  
 CC protein lacking a transmembrane (TM) domain. This sequence is generated  
 CC by the deletion of the C-terminal TM domain of human Asp 2(a) protein.  
 XX  
 SQ Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 2, 4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EMDPEPERGGRGSEVEMVNDLRKSGCGYVEMTVGSPPTLNILVDTGSSNFAGVGAAP 60  
 DB 46 etdeepepgrgrgsefvemvndlrksgcggyvemtvgspptlnilvdtgssnfavgap 105  
 QY 61 HPFLHRYVQROLSTSTYRDLRKGVVVPYTOGKEGELCTDLVSTPHGNVYVRANIAITE 120  
 DB 106 hpflhryvqrolststtyrdlrkgyvvpctogkewelctdlvstphgnvvranaiaite 165

QY 121 SDKFINSNMWEGILGLAYAEIARPPDSLPEFFDSLKQTHVPLNLSQLCGAGPPLNOS 180  
 DB 166 sdkffingsnmwegilglayaelarppdslepffdslkvqthvplnlsqcgagfplngs 225  
 QY 181 EYLAIVGCGMIITGGIDHSLYTGSLMYTTRIRRWYEVIIIVVEINGODLKMCKEYANDK 240  
 DB 226 eylavvgcgmiitggidhsllytgslmyttrirrwyeviivveingodlkmckeyandyk 285  
 QY 241 SYVDSGTTNLRLEPKVFPEAAVKSIIKAASSTKFPDGFPLGPOLYCQAGTPTMNIIPVIS 300  
 DB 286 syvdsqtnlrlpkvfpfaavksikaasstkfpdgfplgqylvcwqagttptmniipvis 345  
 QY 301 LYLMGEVTVNOSFRITILPQYLRLPEVEDVATISQDDCYKFAISQSSGTVMGAVIMEGFYV 360  
 DB 346 lylmgevtnosfrtilpqqylrpvedvatsqddcykfalsqssgtvmgavimegfyvv 405  
 QY 361 FDRARRKRGFAVSACHYDEFRTAAVEGPFYTLMEQCGYNIPTDE 407  
 DB 406 fdrarkrgrgfavsachydefrtaavegpfvtlmedcgyniptde 452

RESULT 4  
 ID AAE06872 standard; Protein: 453 AA.  
 XX  
 AC AAE06872;  
 XX  
 DT 23-OCT-2001 (first entry)  
 XX  
 DE Human-Asp2(a) deltaTM protein.  
 XX  
 KW Human: aspartyl protease 2a; Asp 2a: beta-amyloid precursor protein; APP;  
 KW beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;  
 KW neurofibrillary tangle; neuronal loss; amyloid-beta peptide; neurotrophic;  
 KW neuroprotective; antisense therapy; Asp2(a) deltaTM protein;  
 XX gene therapy.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 214 /note="Encoded by CAC"  
 XX  
 PN WO200150829-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 09-MAY-2001; 2001MO-IB00799.  
 XX  
 PR 09-MAY-2001; 2001MO-IB00799.  
 XX  
 PA (BIEN/) BIENKOWSKI M J.  
 PA (GURN/) GURNEY M E.  
 PA (HEIN/) HEINRIKSON R L.  
 PA (PARO/) PARODI L A.  
 PA (YANR/) YAN R.  
 XX  
 PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
 XX  
 DR WPI: 2001-483072/52.  
 DR N-PSDB; AAD13034.  
 XX  
 PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
 PT activity -  
 XX  
 PS Claim 149; Fig 11; 185pp; English.  
 XX  
 CC The invention relates to human aspartyl proteases (hu-Asp), beta-amyloid  
 CC precursor protein (APP) isoforms and their corresponding DNA molecules.

CC Human aspartyl proteases can act as beta-secretase proteases useful for  
 CC treating Alzheimer's disease. APP isoforms are useful for identifying  
 CC modulators of amyloid-beta peptide production, for use in designing  
 CC therapeutics for the treatment and prevention of Alzheimer's disease,  
 CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis  
 CC and neuronal loss. APP isoforms are also used in methods for identifying  
 CC inhibitors and modulators of human Asp2 activity. The invention relates  
 CC to a method for identifying agents that modulate the activity of human  
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
 CC as a means to screen in cellular assays for the inhibitors of beta- and  
 CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
 CC polymerase chain reactions (PCR). The probes are useful for detecting  
 CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
 CC blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a)  
 CC deltaTM protein which is obtained by the deletion of transmembrane domain  
 CC at the C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase  
 CC activity.

XX Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ETDEPEEPGRGSEFVEMVDNLRGKSGGYVEMTVGSPQTLNLTVDGSSNFAVGAAP 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 46 etdeepeepgrgrsfvemdnlrgksggyvemtvgspqtlntlvdgssnfvgaap 105  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 61 HPELHRYTORLSSTYRDLRKGYVPYTGKWEGLGTLVSIIPGPNTVYRANITAE 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 106 hpehlhrytorlststyrdlrkgvypytgkwegelgtlvsiipgpntrvranitaale 165  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 121 SDKFFINGSNMGEIIGLAEIARPDSDLEPFDSLVKOTHPNLEFSLQCGAGPELND 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 166 sdrffingsnmgeiiglaeiaarpdpsdlepfdsllvkvqthvnpnlfsqlcgagpdlngs 225  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 181 EYLAASVGSMTIGIDHSLYTGLWYTPTRREKYEVIIIVAREINODKMKCKENYDK 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 226 eylasvgsmtlgyidhslytglwtprtrekyevliivareinodkmdckekenyk 285  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 241 SYDSGFTNLRLEPKYFEAAVKSIRKASSTKEKPDGFMVLGEOLVCMQAGTTPWNI 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 286 sydsftnlrlpkkyfeaaavksirkaasstkekpdpqfvlvgeqlvcwagtltpwnlfpvls 345  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 301 LYLMGEVYNOSFRITILIPQYLRPVEDVATSDQCYKFAISOSSTGTWVGAVIMEGYV 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 346 lylmgevtnsfrililpqyilrpvedvatstdqcykfalsgstglvmgavimegyfyyv 405  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 361 FDRARRKRGFAVSACHVDEFTAAVEGPVTILDMEDCGYNIPOIDE 407  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 406 fdrarrkrgfavsachvdeftaaavegpvtildmedcgynipptide 452

RESULT 5  
 AAU06616 ID AAU06616 standard; Protein; 453 AA.

XX AAU06616;

XX 24-OCT-2001 (first entry)

DE Human-pro-Asp 2(a) delta TM.

XX Human; Aspartyl protease; beta-secretase; nootropic; ASP2;

KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;

XX amyloid beta; Abeta; Human-pro-Asp 2(a) delta TM, mutant; mutein.

OS Homo sapiens.

XX Synthetic.

FT Key

FT Peptide

Location/Qualifiers  
 1..21

FT Protein /label= signal\_peptide  
 FT 22.453 /label= Mature\_Human\_pro\_Asp-2(a)\_delta\_TM  
 FT Misc-difference 214 /note= "Encoded by CAC"

PN WO200149098-A2.

PD 12-JUL-2001.

PF 09-MAY-2001; 2001WO-1B00798.

PR 09-MAY-2001; 2001WO-1B00798.

PA (BIEN/) BIENKOWSKI M J.

PA (GURNEY/) GURNEY M E.

PA (HEIN/) HEINRIKSON R L.

PA (PARO/) PARODI L A.

PA (YANR/) YAN R.

PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;

DR WPI; 2001-502549/55.

DR N-PSDB; AAS11530.

XX Claim 149; Page 160; 185pp; English.

The invention relates to a purified polypeptide comprising a fragment of mammalian aspartyl protease (Asp2) protein which lacks the Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide and the fragment retain the beta-secretase activity of the mammalian Asp2 protein. The invention also details polynucleotides for the Asp2 protein and vectors expressing them, and a polypeptide (isoform of amyloid protein precursor (APP)) comprising the amino acid sequence of an APP or its fragment containing an APP cleavage site recognizable by a mammalian beta-secretase, and further comprising two lysine residues at the carboxyl terminus of the amino acid sequence of the mammalian APP or APP fragment. Also included in the invention are methods of identifying APP fragment. Also included in the invention are methods of identifying modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are useful for treating Alzheimer's disease. APP is useful in methods for identifying inhibitors or modulators of human Asp2 activity and amyloid-beta (Abeta) peptide production. APP is also useful in designing therapeutics for the treatment or prevention of Alzheimer's disease. APP comprising the APP-SW-beta-secretase peptide sequence (NLDA), which is associated with increased levels of Abeta processing is useful in assays relating the Alzheimer's research. The expression vector is useful for recombinantly expressing APP. Nucleic acids that hybridize to CC Asp oligonucleotides are useful as probes or primers. The probes are useful for detecting Hu-Asp nucleic acids in in vitro assays and in Northern and Southern blots. The present sequence is Human-pro-Asp 2(a) delta TM protein, which lacks the C-terminal transmembrane domain.

XX Sequence 453 AA;

Query Match 100.0%; Score 2156; DB 22; Length 453;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ETDEPEEPGRGSEFVEMVDNLRGKSGGYVEMTVGSPQTLNLTVDGSSNFAVGAAP 60

DB 46 etdeepeepgrgrsfvemdnlrgksggyvemtvgspqtlntlvdgssnfvgaap 105

OY 61 HPELHRYTORLSSTYRDLRKGYVPYTGKWEGLGTLVSIIPGPNTVYRANITAE 120

DB 106 hpehlhrytorlststyrdlrkgvypytgkwegelgtlvsiipgpntrvranitaale 165



QY	121	SKFFIINSNMEGIIIGLVAIEIARPDSDLEPFEDSLVKQIHVPNLSLOLCGAFPLNQS	1801
Db	166	sdkffingsnwegllyglayaelarppdslepfefslkvkcthpnlfaqlcgagflngs	225
QY	181	EVLAVGVGSMIIIGIDHSLYGSILWYPIRREWYEVYIIVRVEINQDILCKDKEYNKK	240
Db	226	evlaevgsmsllggldhsllygstlwybpilrweyevliivrveingdilkmdkeynydk	285
QY	241	SLVDSGTINLRLPKKVFEEPAANKSIRKAASSTEFKDFGWLGEOLCWMQAGTTPNNIPFVIS	300
Db	286	slvdsgtlnlrlpkkvtfaaaksikaasstekfpgdglwlgelqvcwagcttpnifprvis	345
QY	301	LYLMEVINSQSFRIITLPEQQLRPEVEVATSDODCCYFAISOSSTGVGMAVAMEGFYVV	360
Db	346	lylmevngsfriltlpqqllrpvedvacsqddccyfalsgstctgvmgavamegfyvv	405
QY	361	FDRARKRIGFAVSACNHNDEFRRTAAVSGPVTLLDMEDCGNIPQTDJE	407
Db	406	fdrarkrigrfavsachnndeirrtavaesgpfvtllmedccgynipqtdje	452

```

RESULT      6
AA007215   standard; Protein; 453 AA.
ID          AA007215
AC          AA007215;
XX
DT          24-OCT-2001 (first entry)
XX
DE          Human aspartyl protease 2a deltatm (Huasp-2adeltatm).
XX
KW          Human: aspartyl protease 1; Asp-1; nocrotropic; neuroprotective;
KM          aspartyl protease 2; Asp2; amyloid protein precursor; APP;
KV          beta-secretase; Alzheimer's disease; Huasp-2adeltatm.
XX
OS          Homo sapiens.
XX
FH          Key
FT          Peptide
FT          Location/Qualifiers
FT          1..21
FT          /note= "Signal peptide"
FT          22..453
FT          /note= "Mature human aspartyl protease 2a deltatm"
XX
PD          WO200149097-A2.
PN          12-JUL-2001.
XX
PF          09-MAY-2001; 2001WO-IB00797.
PR          09-MAY-2001; 2001WO-IB00797.
XX
PA          (BIEN/) BIENKOWSKI M J.
PA          (GURNEJ/) GURNEY M E.
PA          (HEINR/) HEINRIKSON R L.
PA          (PARODI/) PARODI L A.
PA          (YANR/) YAN R.
XX
PI          Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;
XX          WPI: 2001-502548/55.
DR          N-PSDB: AAS11715.
XX
PT          Novel purified polypeptide comprising fragment of mammalian aspartyl
PT          protease 2, lacking Asp2 transmembrane domain and retaining beta
PT          secretase activity of Asp2 useful for identifying inhibitors of Asp2
PT          activity -
XX
PS          Claim 149; Fig 11; 185pp; English.
CC

```

protein. Also included is an isoform of amyloid protein precursor (APP) comprising the amino acid sequence of a APP or its fragment containing an APP cleavage site recognisable by a mammalian beta-secretase, and further comprising two lysine residues at the carboxyl terminus of the amino acid sequence of the mammalian APP or APP fragment. The polypeptides are used for assaying for modulators of beta-secretase activity; identifying agents that inhibit the APP processing activity of human Asp2 aspartyl protease (hu-Asp2); identifying agents that modulate the activity of Asp2; and for reducing cellular production of amyloid beta (A $\beta$ et) from APP. Agents identified by the above methods are useful for treating Alzheimer's disease, and for identifying modulators of amyloid-beta (A $\beta$ et) peptide production, for use in designing therapeutics for the treatment or prevention of Alzheimer's disease. Probes and primers derived from APP nucleic acid sequences are useful for detecting Hu-Asp nucleic acids in *in vitro* assays and Northern and Southern blots. The present sequence represents the amino acid sequence of human Asp2a delta TM construct which lacks the transmembrane domain. This construct was used for bacterial expression and purification of human Asp2a.

Query Match	100.0%	Score 2156	DB 22	Length 453
Best Local Similarity	100.0%	Pred. No. 2,4e-210		
Matches 407	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	ETDEPEEPGRGSGFEVEMVDNLRGKSGGQYVEMTVGSPPOFTNLITVPGSSNFAVGAAP	60		
Db	etdeepeepgrgrstfemvndnlrgksqggyvemyvsgppqlnlivdtgssnfavgaap	105		
QY 46	etdeepeepgrgrstfemvndnlrgksqggyvemyvsgppqlnlivdtgssnfavgaap	105		
Db	etdeepeepgrgrstfemvndnlrgksqggyvemyvsgppqlnlivdtgssnfavgaap	105		
QY 61	HPFLHRYORQLSSVYRDLRKGVVYPTYGKWEGLGTDIVSIPHGPNVTYRANIAAIRE	120		
Db	hpflhryqgrqlssteyrdlrgkyvpyptgkwegelgtdlvsiphgpnvtrantlaalte	165		
QY 121	SKKEFLNSNMGGIIGLVAETARDDSLPEPFDSLTKOTHPNFSIQLGAGFPLNOS	180		
Db	skkfllnsnmegllglvyaetaaraddslpfpdlsivkqthpnlfsiqagfplnqs	225		
QY 166	sdkfllnsnmegllglvyaetaaraddslpfpdlsivkqthpnlfsiqagfplnqs	225		
Db	sdkfllnsnmegllglvyaetaaraddslpfpdlsivkqthpnlfsiqagfplnqs	225		
QY 181	EVLASVGGSMITGGIDHSLYTGSLWYTEIRREMYEVLIVREINGQDLKMDCKEYNDK	240		
Db	evlasvgsmsitggidhsltygslwtyeirremyevlrvreingqdlkmdckeyndk	285		
QY 241	SVVDSGTNLRLPKKVFEEAAKSTRAASSTKEFPDGEWLGEOVLCWQAGTTPMNTFPVIS	300		
Db	svvdsgrtnlrlpkkvfeaaakstraasstekfpdgewlgvowaggttppmnlfpvis	345		
QY 301	LYLMEVYNOSFRITILPQOYLPRVEDVATGDDCCYKRAISQSSNGTYMGAVIMGEFVYV	360		
Db	lylmevynofritilpqoylprvedatvsqddcykrtaisqssngtyvmgavimegfyyv	405		
QY 361	FDRAARRIGFAVSACHVDEFRITAAVEGPFVTLDMEDDGYGINPQIDE	407		
Db	fdraarkrigfavsachvdefritaavegpfvtdlmededgyinipqide	452		
RESULT 7				
AAE02594				
ID AAE02594 standard; Protein: 453 AA.				
XX AAE02594:				
XX	10-AUG-2001	(first entry)		
DE	Human-Asp-2(a)	delta TM protein.		
XX				
KM	Human: alpha-secretase; amyloid precursor protein; APP; therapy;			
KM	Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp2a;			
XX	beta-secretase; Asp-2a	delta TM.		
OS	Homo sapiens.			
OS	Synthetic.			
XX				

Key Location/Qualifiers  
Misc-difference 214  
/note= "Encoded by CAC"

MO200123533-A2.

05-APR-2001.

22-SEP-2000; 2000WO-US26080.

23-SEP-1999; 99US-0155493.

23-SEP-1999; 99WO-US20881.

13-OCT-1999; 99US-0416901.

06-DEC-1999; 99US-0169232.

(PHAA ) PHARMACIA & UPJOHN CO.

Gurney M, Bienkowski MJ;

WPI; 2001-290516/30.

N-PSDB; AAD06752.

Enzymes that cleave the alpha-secretase site of the amyloid precursor protein, useful for the treatment of Alzheimer's disease -

Example 10; Fig 11; 189pp; English.

The present invention relates to enzymes for cleaving the alpha-secretase site of the amyloid precursor protein (APP) and methods of identifying those enzymes. The methods may be used to identify enzymes that may be used to cleave the alpha-secretase cleavage site of the APP protein. The enzymes may be used to treat or modulate the progress of Alzheimer's disease. The present sequence is human Aspartyl protease 2a (Asp-2a) deltatm protein which is obtained by deleting its transmembrane domain. This sequence has beta-secretase protease activity.

Sequence 453 AA:

Query Match 100.0%; Score 2156; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 2,4e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ENDEEPEEGRGRSFVEMVDNLRGSGGQYVEMTVGSPPTLNILVDTGSSNFAVGAAP 60

46 etdeepeepgrgrsfvemdnlrgsggqyvyemtvgsppqlnlldvsgssnlavgaap 105

61 HPELHRYRQRLSSYRDLRKGVVYPYTGQKMEGELGTDLVSIIPHGPNTVRANIATTE 120

106 hpehlhryrqrlssyrdlrkgvvyptgqkwegelgtdlvsiphgpnvtraniaalte 165

121 SDRFFINGSNWEGILGLAEIARPPDLSLEPPFDSLVRQTHVNFSLQCGAGFPLNQS 180

166 sdrffingsnwegilglayaelarppdlslepfdsivkqthvnlfsiqcagafplnqs 225

181 EYLASVSGSMITIGIDHSIYTGSLWYTPIRREYVEVIVRVEINQODLKMCKEKNYDK 240

226 eylasvsgsmiitgidslytgslywtpirreyvevivrveinngdtkmckekeyndk 285

286 slyasvsgsmiitgidslytgslywtpirreyvevivrveinngdtkmckekeyndk 345

301 LYLMGEVYTNOSFRITLIPQYLIRPVEDVATSDQCYKFAISQSSSTGVNMGAVIMEGYVY 360

346 lylmgevntnosfrtllipqylirpvedvatsdqcykfaalsgststgvnmgavimegyvy 405

361 FDRARRKIGFAVSACHVDEFRFAAVGPEVTLDMEDCGYNIPOQTD 407

406 fdarrkigfavsachvdefrfaavgpvtlmedcgynipqtd 452

AAB07897  
ID AAB07897 standard; Protein; 456 AA.

AAB07897;

14-NOV-2000 (first entry)

Active enzyme portion of human beta-secretase enzyme.

Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide; amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

inhibitor; ss.

Homo sapiens.

WO200047618-A2.

17-AUG-2000.

10-FEB-2000; 2000WO-US03819.

10-FEB-1999; 99US-0119571.

15-JUN-1999; 99US-0139172.

(ELAN-) ELAN PHARM INC.

Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

WPI; 2000-533011/48.

Purified beta-secretase protein used in assays to discover inhibitors which can be used for the treatment of amyloidogenic diseases e.g.

Alzheimer's disease -

Claim 24; Fig 2b; 121pp; English.

The specification describes a beta-secretase enzyme. The enzyme cleaves beta-amyloid precursor protein to produce beta-amyloid peptide. This

enzyme is therefore implicated in the production of amyloid plaque components which accumulate in the brains of individuals afflicted with

Alzheimer's disease. Inhibitors of beta-secretase are administered to

a mammalian subject e.g. with Alzheimer's disease or improve cognitive

disease-like pathology to test if they maintain or improve cognitive

ability or reduce the plaque burden. The compounds are used for the

treatment of amyloidogenic diseases e.g. Alzheimer's disease. The

present sequence represents the active enzyme portion of human

beta-secretase enzyme.

Sequence 456 AA:

Query Match 100.0%; Score 2156; DB 21; Length 456;  
Best Local Similarity 100.0%; Pred. No. 2,4e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ENDEEPEEGRGRSFVEMVDNLRGSGGQYVEMTVGSPPTLNILVDTGSSNFAVGAAP 60

1 etdeepeepgrgrsfvemdnlrgsggqyvyemtvgsppqlnlldvsgssnlavgaap 60

61 HPELHRYRQRLSSYRDLRKGVVYPYTGQKMEGELGTDLVSIIPHGPNTVRANIATTE 120

61 hpehlhryrqrlssyrdlrkgvvyptgqkwegelgtdlvsiphgpnvtraniaalte 120

121 SDRFFINGSNWEGILGLAEIARPPDLSLEPPFDSLVRQTHVNFSLQCGAGFPLNQS 180

121 sdrffingsnwegilglayaelarppdlslepfdsivkqthvnlfsiqcagafplnqs 180

181 EYLASVSGSMITIGIDHSIYTGSLWYTPIRREYVEVIVRVEINQODLKMCKEKNYDK 240

181 eylasvsgsmiitgidslytgslywtpirreyvevivrveinngdtkmckekeyndk 240

241 SIVDSGTNTLRPKKVFEEAFAVSIKAASTKEKFPDGFMLGELVLCWQAGTTPWNIFFPVIS 300

Db 241 sivdsqfntlnrlpkkvfeaaavksikaasstekfpdgfwagdlvwqagatlpwnlfpvis 300  
|  
Qy 301 LYLMGEVTNNSFKITLLIPQOYLRFVEDVATSDDCKFAISQSSGTVMGAVIMEGFYV 360  
|  
Db 301 LYIMGEVTNNSFKITLLIPQOYLRFVEDVATSDDCKFAISQSSGTVMGAVIMEGFYV 360  
|  
Qy 361 FDSARRKRGFAVSAACHVHDEFFRTAAVEGPFVTLDMEDCGYNLPQIDE 407  
|  
Db 361 fdrarrtrigfavsachvhdftrtaavegpfvrltmedcgynlpqtde 407  
|  
RESULT 9  
AAT88439  
ID AAY88439 standard; Protein; 459 AA.  
XX  
XX AAY88439;  
XX  
XX  
DT 03-AUG-2000 (first entry)  
XX  
DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.  
XX  
XX Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;  
KW Alzheimer's disease; beta secretase site.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200017369-A2.  
PN  
PD 30-MAR-2000.  
XX  
XX 23-SEP-1999; 99MO-US20881.  
PF  
XX  
XX 24-SEP-1998; 98US-O101594.  
PR  
XX  
XX (PHAA ) PHARMACIA & UPJOHN CO.  
PA  
PI Gurney ME, Bienkowski MJ, Heintzson RL, Parodi LA, Yan R;  
PI  
XX  
DR WPI: 2000-303209/26.  
DR  
XX N-PSDB: AAI15689.  
XX  
PT New enzyme designated human aspartase useful in research into  
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at  
XX the beta secretase site to produce amyloid beta peptide  
XX  
PS Example 10; Page 173-176; 163pp; English.  
XX  
CC This sequence represents a modified human aspartyl protease 2 (Asp2)  
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal  
CC transmembrane domain deleted. The invention relates to a protease  
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid  
CC precursor protein (APP). The protease contains a sequence encoding the  
CC amyloid acid sequence DNG and a sequence encoding D56 or DNG separated by  
CC 100-300 amino acids. When mutated the APP gene causes an autosomal  
CC dominant form of Alzheimer's disease. The invention includes a  
CC membrane and have a single C-terminal transmembrane domain. Proteolytic  
CC processing of APP produces the amyloid beta protein, which is possibly  
CC very important in Alzheimer's disease. APP localises to the cell surface  
CC nucleotide sequence encoding the protease. The invention includes a  
CC nucleotide sequence, and a cell line comprising the vector. Methods for  
CC screening for inhibitors of beta secretase activity are also given in the  
CC invention. The human aspartase protein and nucleotide sequences and the  
CC methods for identifying inhibitors of the protease, are useful in the  
XX treatment of and research in to Alzheimer's disease.  
XX  
XX Sequence 459 AA;  
XX

Query Match	100.0%;	Score 2156;	DB 21;	Length 459;
Best Local Similarity	100.0%;	Pred. No. 2.5e-210;		
Matches 407; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	ETDEEPEPPGRCSFPEMDNLRGSGGCIYEMTVGSPPTOTLILVDGSSNFAYGAAP	60
Db	46	etdeepeepgrcsfpeymdnlrqksqggyvemtvsppotlilvdcgssnfavgaaP	100
QY	61	HPFLHRRYQBOLSSTRDYLPRKGVYPPYQGWKEGELGDLVSIPIHGNNTVRANLAAATE	120
Db	106	hpflhryyqrqlastcyrdlrlrkgyvpyrcqgkwegejldlvsipbnpvtvraanaahte	160
QY	121	SDKFFINGSNWBEILCLAYAEFLARPDDSLPEFPDSLVKQTHVBNPLSLDLCAGFPLNOS	180
Db	166	sdckfflngsmwewilg1ayaeelarppdslepffqslqkqchvnpnltslqcgafp1nqs	220
QY	181	EVLASVSGSMIIIGIDHSILYTGSLMYTPRIREWEYEVIIYRVLEINODLKMOCKEYNDK	240
Db	226	evlasvsgsmliigldhslytgslsylycp1rlreweyev1lyrvleingqdlkmocekeynydk	280
QY	241	SYDSGTNNLRPKVFEAAVKSIAASTSEKFPDGCWMLCEOLVYCAQAGTPPMN1FEPIS	300
Db	286	sydsqgtntnlrpkvkfeaaavksikaassteekfpdgfw1geq1vcwgagsttprn1fpris	340
QY	301	LYLMGETNOSRRITLILQYILRPVEDYVANSODDCYKFAISGSSTGTWYAGVIMEGFYV	360
Db	346	lylmgetnqstftrililpqyilrpvedatalsqddcykfalsqscgtvmgavimegfyyv	400
QY	361	FDBARRKIGFAVSACHYHDERFRAAVGSPFTLLDMEDCGN1POND 407	
Db	406	fdrarrk1gfvavachynderftraavagspftllamedcgyn1pndte 452	

RESULT	10
AAE10643	
ID	AAE10643 standard; Protein; 459 AA.
AC	AAE10643;
XX	
DT	10-DEC-2001 (first entry)
XX	
DE	Human-Asp 2(a) protein with (His)6 tag and lacking TM domain.
XX	
XX	Human: aspartyl protease 2a; Asp2a; amyloid precursor protein; APP;
KW	Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
XX	amyloid plaque; neuronal loss; proteolytic; neurotrophic; neuroprotective
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..21
FT	/label= Signal_peptide
FT	Protein
FT	22..459
FT	/note= "Mature Human-Asp2(a) delatam (His)6 protein"
FT	Misc-difference 214
FT	/note= "Encoded by CAC"
FT	Misc-difference 454
FT	/note= "Encoded by CAG"
FT	Misc-difference 435
FT	/note= "Encoded by CAG"
FT	Misc-difference 456
FT	/note= "Encoded by CAG"
FT	Misc-difference 457
FT	/note= "Encoded by CAG"
FT	Misc-difference 458
FT	/note= "Encoded by CAG"
FT	Misc-difference 459
FT	/note= "Encoded by CAG"
FT	Misc-difference 459
FT	/note= "Encoded by CAG"
XX	
PN	GB2357767-A.
XX	
PD	04-JUL-2001.
XX	
PF	22-SEP-2000; 2000GB-0023315.
XX	
PR	23-SEP-1999; 99US-0155493.



modulators of amyloid-beta peptide production, for use in designing  
 CC therapeutics for the treatment and prevention of Alzheimer's disease,  
 CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis  
 CC and neuronal loss. APP isoforms are also used in methods for identifying  
 CC inhibitors and modulators of human Asp2 activity. The invention relates  
 CC to a method for identifying agents that modulate the activity of human  
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used  
 CC as a means to screen in cellular assays for the inhibitors of beta- and  
 CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in  
 CC polymerase chain reactions (PCR). The probes are useful for detecting  
 CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern  
 CC blots. The present sequence is Human aspartyl protease 2a (Hu-Asp2a)  
 CC deltaTM (His)6 protein which is obtained by deletion of C-terminal  
 CC transmembrane domain and addition of a hexa-Histidine tag at the  
 CC C-terminal end of Hu-Asp2a. Human Asp2a has beta-secretase activity.  
 CC  
 XX  
 SQ Sequence 459 AA:

Query Match 100.0%; Score 2156; DB 22; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ETDEEPEEPRGRGSFVEMVDNLGRKSGGGYVEMTVGSPPTNTIIVDGGSSNFAYGAAP 60  
 46 etdeepeepgrgsfvevmdnlrgksgggyvemtvgspptntllvdcgssnfaygaap 105  
 61 HPLHRYQRQLSSTYRDLRKGYVYPYTOGKWEGLGTLVSLPHGPNYTVRANIAITE 120  
 106 hplhryqrqlsstyrdlrkgyvpytoggkweglgtlvsiphgpnvtraniaite 165  
 121 SDFEFGNSWEGTILGAAVAETARPDDSLPEPDSLVKQTHVNFSLDLCAGAFPLNOS 180  
 166 sdfefngswegtllgaavaeetarppdslepdslvkqthvnfnslldcagafplnos 225  
 181 EVLASVGSMTIGIDHSLSYTGSLMWTPIRREYVEVILVIREINSGODKMKCKENYOK 240  
 226 evlasvgsmtigidslsytgslmwtpirreyvevillvireinsgodkmkckenynok 285  
 241 SYDSTTNLRPLPKKFEAAVSKRAASSTEFKPDGFWLGEOLVCMQAGTTPWNIFFPVIS 300  
 286 sydsttnlrplpkkfeaaavskraasstefkpdgfwlgeolvcmqagtppwniffpvls 345  
 301 LYLMEVETNQSFRTITLPQOYLRPVEDVATSDDDCKKFAISQSSGTVMGAVIMEGFYV 360  
 346 lylmevetnqsfritlppqylrpedvatsddckkfaissgstvmgavimegfyv 405  
 361 FRARRKRIFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTD 407  
 406 fdrarkrifavsachvhdefrtaavegpfvtlmedcgynipqtd 452

RESULT 12  
 AAU06617  
 ID AAU06617 standard: Protein: 459 AA.

AC AAU06617;

DT 24-OCT-2001 (first entry)

DE Human-pro-Asp 2(a) delta TM (His)6.

KM Human: Aspartyl protease; beta-secretase; nootropic; Asp2;

KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;

OS amyloid-beta; Abeta; Human-pro-Asp 2(a) delta TM (His)6; mutant; mutlein.

OS Homo sapiens.

XX Synthetic.

XX Key

FT Peptide

FT Protein

Location/Qualifiers  
 1..21  
 /label= Signal\_peptide  
 22..459  
 Protein

FT /label= Mature\_Human\_Pro-Asp-2(a)\_delta-TM\_(His)6  
 FT Misc-difference 214 /note= "Encoded by CAC"  
 FT Misc-difference 454..459 /note= "Encoded by CAGCAGCAGCAGCAGCAGC"  
 FT Region 454..459 /label= His\_tag  
 FT /note= "Nickel binding region to aid purification"  
 PN WO200149098-A2.  
 PD 12-JUL-2001.  
 XX  
 PD 09-MAY-2001; 2001WO-1B00798.  
 PF 09-MAY-2001; 2001WO-1B00798.  
 XX  
 PR 09-MAY-2001; 2001WO-1B00798.  
 XX  
 PA (BIEN/) BIENKOWSKI M J.  
 PA (GURN/) GURNEY M E.  
 PA (HEIN/) HEINRIKSON R L.  
 PA (PARO/) PARODI L A.  
 PA (YANR/) YAN R.

PI Bienkowski MJ, Gurney ME, Heinrichson RL, Parodi LA, Yan R;  
 DR WPI: 2001-502549/55.  
 DR N-PSDB: AAS11531.

PT Novel purified polypeptide comprising fragment of mammalian aspartyl  
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta  
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2  
 PT activity  
 PS  
 XX Claim 149; Fig 12; 185pp; English.

The invention relates to a purified polypeptide comprising a fragment of  
 CC mammalian aspartyl protease (Asp)2 protein which lacks the Asp2  
 CC transmembrane domain and the Asp2 protein, and where the polypeptide and  
 CC the fragment retain the beta-secretase activity of the mammalian Asp2  
 CC protein. The invention also details polynucleotides for the Asp  
 CC proteins and vectors expressing them, and a polypeptide (isoform of  
 CC amyloid protein precursor (APP)) comprising the amino acid sequence of an  
 CC APP or its fragment containing an APP cleavage site recognizable by a  
 CC mammalian beta-secretase, and further comprising two lysine residues at  
 CC the carboxyl terminus of the amino acid sequence of the mammalian APP or  
 CC APP fragment. Also included in the invention are methods of identifying  
 CC modulators or inhibitors of Asp2. Modulators and inhibitors of Asp2 are  
 CC useful for treating Alzheimer's disease. APP is useful in methods for  
 CC identifying inhibitors or modulators of human Asp2 activity and  
 CC amyloid-beta (Abeta) peptide production. APP is also useful in designing  
 CC therapeutics for the treatment or prevention of Alzheimer's disease.  
 CC APP comprising the App-Sw-beta-secretase peptide sequence (NDA), which  
 CC is associated with increased levels of Abeta processing is useful in  
 CC assays relating the Alzheimer's research. The expression vector is useful  
 CC for recombinantly expressing APP. Nucleic acids that hybridize to  
 CC Asp oligonucleotides are useful as probes or primers. The probes are  
 CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in  
 CC Northern and Southern blots. The present sequence is Human-pro-  
 CC Asp 2(a) delta TM (His)6 protein, which lacks the C-terminal  
 CC transmembrane domain and has a His tag to aid purification.  
 XX

SQ Sequence 459 AA:

Query Match 100.0%; Score 2156; DB 22; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ETDEEPEEPRGRGSFVEMVDNLGRKSGGGYVEMTVGSPPTNTIIVDGGSSNFAYGAAP 60  
 46 etdeepeepgrgsfvevmdnlrgksgggyvemtvgspptntllvdcgssnfaygaap 105  
 61 HPLHRYQRQLSSTYRDLRKGYVYPYTOGKWEGLGTLVSLPHGPNYTVRANIAITE 120



ID AAE02595 standard; Protein: 459 AA.  
XX  
AC AAE02595;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Human-Asp-2(a) delcATM (His)6 protein.  
XX  
KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;  
KW Alzheimer's disease; antialzheimer's; aspartyl protease 2a; ASP2a;  
KW beta-secretase; Asp-2a delta TM; histidine tag; mutant; muten.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 214  
FT /note= "Encoded by CAC"  
FT Misc-difference 454  
FT /note= "Encoded by CAG"  
FT Misc-difference 435  
FT /note= "Encoded by CAG"  
FT Misc-difference 456  
FT /note= "Encoded by CAG"  
FT Misc-difference 457  
FT /note= "Encoded by CAG"  
FT Misc-difference 458  
FT /note= "Encoded by CAG"  
FT Misc-difference 459  
FT /note= "Encoded by CAG"  
FT Misc-difference /note= "Encoded by CAG"  
XX  
PN WO200123533-A2.  
XX  
PD 05-APR-2001.  
XX  
PE 22-SEP-2000; 2000WO-US26080.  
XX  
PR 23-SEP-1999; 99US-0155493.  
PR 23-SEP-1999; 99WO-US20881.  
PR 13-OCT-1999; 99US-0416901.  
PR 06-DEC-1999; 99US-0169232.  
XX  
PA (PHMA ) PHARMACIA & UPJOHN CO.  
XX  
PI Gurney M, Blenkowski MJ;  
XX  
DR MPI: 2001-290516/30.  
DR N-PSDB: AAD06753.  
XX  
PT Enzymes that cleave the alpha-secretase site of the amyloid precursor  
PT protein, useful for the treatment of Alzheimer's disease -  
XX  
PS Example 10; Fig 12; 189pp; English.  
XX  
CC The present invention relates to enzymes for cleaving the alpha-  
CC secretase site of the amyloid precursor protein (APP) and methods of  
CC identifying those enzymes. The methods may be used to identify enzymes  
CC that may be used to cleave the alpha-secretase cleavage site of the APP  
CC protein. The enzymes may be used to treat or modulate the progress of  
CC Alzheimer's disease. The present sequence is human Aspartyl protease 2a  
CC (Aap 2a) delcATM (His)6 protein which is obtained by deleting the  
CC transmembrane domain and adding a histidine tag at the C-terminal end.  
CC This sequence has beta-secretase protease activity.  
XX  
SQ Sequence 459 AA;

Query Match 100.0%; Score 2156; DB 22; Length 459;  
Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETDEPEEPGRGSPVEVDLRRKSGOGYVETVGSPPQTLIIIVDTGSSNFAVGAAP 60  
|||||

Db 46 etdepeepgrtgsfemvndljgksqggyvemtvsgppqtlnllvdtgssnfavgaap 105  
QY 61 HPELHRYGROLSTSRDLRKGVYVETGKMEGELGPTDVGTPHGPNTVANAIAATE 120  
Db 106 hpfclhrygrqlastyrdrirkgyvpytqgkwegelqdlvslphgpnvtvraanaaite 165  
QY 121 SDKFFINGSNMBSITGLAVAEIARPDSDLPEPDSLVKQTHPNFLSLQCGAGFPLNS 180  
Db 166 sdrffingsnwegllglayaelarpdsslepfdsllvqthvnpnlfsqldcgagfplns 225  
QY 181 EYLAVSGSMIIGTIDHSYTGSLWTPIRREWYEVIIIVRVEINGODIKMDCKEYNDK 240  
Db 226 evlasvgsmltggldhsllytqslwytprremyevllivrelngqdlkmdckeyndk 285  
QY 241 STVDSGTTNLRPKVFEAAVKSIRKASTSEKFPDGFMLGEOYQWOGTTPWNTFPIYS 300  
Db 286 slvdsqgtlnlrpkvfeavksikaastekfpdgfalgelvwqgqtlpwnlfpiys 345  
QY 301 LYMGSEVTWQSFRTITLROQYLRPYEDVATSDDCYKFAISQSTGYVGAVIMEGFPYV 360  
Db 346 lylmgsevtwqsfritlrlpqylrpvedvatsqdcykfaisqstgyvmgavimegfyv 405  
QY 361 FDRARRIGFAVSAACHVDEPFAVEGPPFTLMDGCGYNTPTDE 407  
Db 406 fdrarkrigfavsachvdefrtaavegpfvltmdcgyniptde 452

RESULT 15  
AAB07898  
ID AAB07898 standard; Protein: 460 AA.  
XX  
AC AAB07898;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE Amino acid sequence of a human beta-secretase enzyme fragment.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
OS  
XX  
PN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PE 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Stlnh S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR MPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
PS Claim 55; Fig 3A; 121pp; English.  
XX

The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the

CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a human beta-secretase enzyme fragment.  
XX  
SQ Sequence 460 AA:

Query Match 100.0%; Score 2156; DB 21; Length 460;  
Best Local Similarity 100.0%; Pred. No. 2.5e-210;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEPEEPGRSGSEVEMVDNLKSKSQGYVEMTSGSPPTNLILVDTGSSNFAYGAAP 60  
DB 46 etdepeepgrsgsevemvndnlrksqgyvemtsgspptnlilvdtgssnfaygaap 105  
QY 61 HPELHRYRQQLSTYRDLRKGVVVPYTOGKMEGELGTDLVSIPHGNVTVRANIAATE 120  
DB 106 hpeLhryrqqlstYrDLrKgvVvPYtOGkMEgELgTDlVsIpHgNvTVrAnIAaTe 165  
QY 121 SDKFFINGSMNEGILGLAYAEIARPPDLSLEPFSDLYKQTHVPMLESLQCGAGFPLNOS 180  
DB 166 sdkffingsmnegilglayaeIarPpDdlsLEpFsdLyKqThVpMlESlQcGagfPlnOs 225  
QY 181 EVLASVGSMAIIGGIDHSLYTGSLWYPIRREWYEVITVREINGODLKMCKEYNDK 240  
DB 226 evlasvgsmAiIGgIDhSLyTgSLwYpIRrEWyEVITvREInGODlKMcKEyNdK 285  
QY 241 STVDSGTNLRLPKKVFEAAVAKSIRAASTTEKFPDGFMLGEOLVCWQAGTTPWNIFFYIS 300  
DB 286 stvdsgtNLRlPKKvFEAAvAKsIRAAStTEkFPdGFmLGEOLvCWQAGtTPwNIffYis 345  
QY 301 IYLMGEVTNQSFRITILPQOYLRPEVDVATSDDDCYKFAISQSGTGYMGAVIMEGFYV 360  
DB 346 IylmgevTnQsFRITilPqOyLRpEvdVATsDDdCYkFAISqSGtGYmGAVImEGfYv 405  
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGFVTLDMDCGYNIPQTD 407  
DB 406 fdarrrRiGfAvsAchVhDefRtAAvEGfVtLDMdCGyNIpQtDe 452

Search completed: August 7, 2002, 09:16:15  
Job time: 160 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:14:48 ; Search time 22.58 seconds

(without alignments)  
440.266 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 1 EIDEEPEEGRGKSGVEWVD.....GPEVTLMEDCGYNIPQDE 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2150	99.7	501	US-09-009-191-2	Sequence 2, Appli
2	2073	96.2	774	US-09-009-191-4	Sequence 4, Appli
3	1132	52.5	518	US-08-999-723-2	Sequence 2, Appli
4	1132	52.5	518	US-09-434-427-2	Sequence 2, Appli
5	1116.5	51.8	514	US-09-717-432-2	Sequence 2, Appli
6	1116.5	51.8	514	US-09-912-484-2	Sequence 2, Appli
7	299.5	13.9	396	US-08-208-007A-13	Sequence 13, Appli
8	299.5	13.9	396	US-09-032-523-9	Sequence 9, Appli
9	293.5	13.7	412	US-08-208-007A-12	Sequence 12, Appli
10	293.5	13.7	412	US-08-974-691-4	Sequence 4, Appli
11	279.5	13.0	458	US-09-640-305-6	Sequence 6, Appli
12	273.5	12.7	409	US-08-846-021A-8	Sequence 8, Appli
13	273.5	12.7	409	US-08-846-021A-8	Sequence 8, Appli
14	273.5	12.7	427	US-08-360-673-6	Sequence 6, Appli
15	270	12.5	410	US-08-245-756-2	Sequence 2, Appli
16	270	12.5	410	US-08-441-751-2	Sequence 2, Appli
17	270	12.5	410	US-08-441-751-2	Sequence 2, Appli
18	270	12.5	410	US-08-441-751-2	Sequence 2, Appli
19	270	12.5	410	US-08-441-751-2	Sequence 2, Appli
20	250	11.6	349	PCT-US92-02521-2	Sequence 2, Appli
21	227	10.5	398	US-09-032-523-3	Sequence 3, Appli
22	227	10.5	398	US-08-328-314-2	Sequence 2, Appli
23	212	9.8	419	US-08-974-691-3	Sequence 3, Appli
24	210	9.7	397	US-09-079-415-2	Sequence 2, Appli
25	196.5	9.1	430	US-08-535-237-2	Sequence 2, Appli
26	195.5	9.1	427	US-07-958-222A-2	Sequence 2, Appli
27	193	9.0	330	US-08-115-753-1	Sequence 1, Appli

28	193	9.0	419	US-08-115-753-2	Sequence 2, Appli
29	193	9.0	419	US-08-115-753-3	Sequence 3, Appli
30	187	8.7	420	US-09-008-271A-4	Sequence 4, Appli
31	187	8.7	420	US-08-974-691-8	Sequence 8, Appli
32	184.5	8.6	395	US-08-723-938-3	Sequence 3, Appli
33	184.5	8.6	395	US-09-080-538-3	Sequence 3, Appli
34	184	8.5	445	US-08-974-691-6	Sequence 6, Appli
35	184	8.5	451	US-08-974-691-2	Sequence 2, Appli
36	149	6.9	437	US-09-353-332-2	Sequence 2, Appli
37	129.5	6.0	140	US-09-211-631-13	Sequence 13, Appli
38	129.5	6.0	140	US-09-265-628-13	Sequence 13, Appli
39	129.5	6.0	140	US-09-001-141-11	Sequence 11, Appli
40	129.5	6.0	140	US-09-001-141-11	Sequence 11, Appli
41	129.5	6.0	140	US-09-532-803-6	Sequence 6, Appli
42	97	4.5	1030	US-09-653-403-14	Sequence 14, Appli
43	95.5	4.4	280	US-09-091-117-2	Sequence 2, Appli
44	89.5	4.2	1097	US-09-160-246-14	Sequence 14, Appli
45	88	4.1	746	US-08-680-326-39	Sequence 39, Appli
				US-08-838-219B-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1  
US-09-009-191-2  
; Sequence 2, Application US/09009191  
; Patent No. 6319689  
; GENERAL INFORMATION:  
; APPLICANT: POWER, DAVID  
; APPLICANT: CHAPMAN, CONRAD  
; APPLICANT: MURPHY, KAY  
; APPLICANT: SMITH, TRUDI  
; TITLE OF INVENTION: ASP2  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: RATNER & PRESTIA  
; STREET: P.O. BOX 980  
; CITY: VALLEY FORGE  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19482  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/009,191  
; FILING DATE: 20-JAN-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UK 9701684.4  
; FILING DATE: 28-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PRESTIA, PAUL F  
; REGISTRATION NUMBER: 23,031  
; REFERENCE/DOCKET NUMBER: GH-70368  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-407-0700  
; TELEFAX: 610-407-0701  
; TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 501 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-009-191-2  
Query Match 99.7%; Score 2150; DB 4; Length 501;  
Best Local Similarity 99.8%; Pred. No. 9,4e-223;

Matches 406; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETDEPEPEPRGRGSEFVEMVNLRGKSGGYVEMTVGSPPTQNLIIIVDTGSSNFVGAAP 60  
 Db 46 ETDEPEPEPRGRGSEFVEMVNLRGKSGGYVEMTVGSPPTQNLIIIVDTGSSNFVGAAP 105

QY 61 HPFLHRYQRLSSTYRDLRKGVVVPYTOGKMEGELGTDIVSIPHGPNVYRANIAITE 120  
 Db 106 HPFLHRYQRLSSTYRDLRKGVVVPYTOGKMEGELGTDIVSIPHGPNVYRANIAITE 165

QY 121 SDFEFINSNMEGILGLAYAEIARPDSDLPEFPDSLVKQTHVPLFSIQLCGAGFPLNOS 180  
 Db 166 SDFEFINSNMEGILGLAYAEIARPDSDLPEFPDSLVKQTHVPLFSIQLCGAGFPLNOS 225

QY 181 EYLASVSGSMIIGIDHSLYTGSIMWTPIRREMYEVIIIVVEINGODLKMCKEYNDK 240  
 Db 226 EYLASVSGSMIIGIDHSLYTGSIMWTPIRREMYEVIIIVVEINGODLKMCKEYNDK 285

QY 241 SIVDSGTTNLRPKKVFEEAAVKSIIKAASSTKFPDGFMLGEOLVCWQAGTTPMNIFFVIS 300  
 Db 286 SIVDSGTTNLRPKKVFEEAAVKSIIKAASSTKFPDGFMLGEOLVCWQAGTTPMNIFFVIS 345

QY 301 IYLMGEVTVNOSRTITLPQOYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV 360  
 Db 346 IYLMGEVTVNOSRTITLPQOYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV 405

QY 361 FDRARRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 407  
 Db 406 FDRARRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 452

RESULT 2  
 US-09-009-191-4  
 ; Sequence 4, Application US/09009191  
 ; Patent No. 6319689  
 ; GENERAL INFORMATION:  
 ; APPLICANT: POWELL, DAVID  
 ; APPLICANT: CHAPMAN, CONRAD  
 ; APPLICANT: MURPHY, KAY  
 ; APPLICANT: SMITH, TRUDI  
 ; TITLE OF INVENTION: ASP2  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: RATNER & PRESTIA  
 ; STREET: P.O. BOX 980  
 ; CITY: VALLEY FORGE  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19482  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/009,191  
 ; FILING DATE: 20-JAN-1998  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: UK 9701684.4  
 ; FILING DATE: 28-JAN-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: PRESTIA, PAUL F  
 ; REGISTRATION NUMBER: 23,031  
 ; REFERENCE/DOCKET NUMBER: GH-70368  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-407-0700  
 ; TELEFAX: 610-407-0701  
 ; TELEX: 846169  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 774 amino acids  
 ; TYPE: amino acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-009-191-4

Query Match 96.2%; Score 2073; DB 4; Length 774;  
 Best Local Similarity 99.2%; Pred. No. 3,8e-214;  
 Matches 392; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 13 GSEFVEVNDNLRGKSGGYVEMTVGSPPTQNLIIIVDTGSSNFVGAAPFLHRYQRL 72  
 Db 1 GSEFVEVNDNLRGKSGGYVEMTVGSPPTQNLIIIVDTGSSNFVGAAPFLHRYQRL 60

QY 73 SSTYRDLRKGVVVPYTOGKMEGELGTDIVSIPHGPNVYRANIAITE SDFEFINSNME 132  
 Db 61 SSTYRDLRKGVVVPYTOGKMEGELGTDIVSIPHGPNVYRANIAITE SDFEFINSNME 120

QY 133 GILGLAYAEIARPDSDLPEFPDSLVKQTHVPLFSIQLCGAGFPLNOS EYLASVSGSMII 192  
 Db 121 GILGLAYAEIARPDSDLPEFPDSLVKQTHVPLFSIQLCGAGFPLNOS EYLASVSGSMII 180

QY 193 GGDHSLYTGSIMWTPIRREMYEVIIIVVEINGODLKMCKEYNDK SIVDSGTTNLR 252  
 Db 181 GGDHSLYTGSIMWTPIRREMYEVIIIVVEINGODLKMCKEYNDK SIVDSGTTNLR 240

QY 253 PKKVFEEAAVKSIIKAASSTKFPDGFMLGEOLVCWQAGTTPMNIFFVIS IYLMGEVTVNOS 312  
 Db 241 PKKVFEEAAVKSIIKAASSTKFPDGFMLGEOLVCWQAGTTPMNIFFVIS IYLMGEVTVNOS 300

QY 313 RITILPQOYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV FDRARRIGFAV 372  
 Db 301 RITILPQOYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV FDRARRIGFAV 360

QY 373 SACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 407  
 Db 361 SACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQDE 395

RESULT 3  
 US-08-999-723-2  
 ; Sequence 2, Application US/08999723A  
 ; Patent No. 6025180  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Powell, David J.  
 ; APPLICANT: Southan, Christopher  
 ; APPLICANT: Chapman, Conrad G.  
 ; APPLICANT: Evans, Joanne R.  
 ; TITLE OF INVENTION: ASP1  
 ; FILE REFERENCE: GH70262  
 ; CURRENT APPLICATION NUMBER: US/08/999,723A  
 ; CURRENT FILING DATE: 1997-10-06  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 518  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-08-999-723-2

Query Match 52.5%; Score 1132; DB 3; Length 518;  
 Best Local Similarity 54.0%; Pred. No. 4.1e-113;  
 Matches 209; Conservative 65; Mismatches 109; Indels 4; Gaps 2;

QY 9 PGRGRGSEFVEMVNLRGKSGGYVEMTVGSPPTQNLIIIVDTGSSNFVGAAPFLHRY 68  
 Db 71 PGRGRGSEFVEMVNLRGKSGGYVEMTVGSPPTQNLIIIVDTGSSNFVGAAPFLHRY 130

QY 69 QRLSSTYRDLRKGVVVPYTOGKMEGELGTDIVSIPHGPNVYRANIAITE SDFEFINSNME 128  
 Db 131 QRLSSTYRDLRKGVVVPYTOGKMEGELGTDIVSIPHGPNVYRANIAITE SDFEFINSNME 190

OY 129 NMWEEIIEIAEELARPDDSLPEPFDSLVTQTHPNPNEFSIOLGAGFPLNQSEVLASVGG 188  
| | | | | : | | | | | : | | | | | : | | | | | : |  
Db 121 IKMNIICGLATATLAKRBSLSLETFPDSLVTQANTPNPNSWMQCGAGLPVAGS--GTNGC 247  
OY 189 SMIIGIDHSLYTGSSLMWTPPIRRMEYYEVIIIVREINGODLKMDCKEKYANDKSIDVGTT 248  
:::|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:  
Db 248 SLVYGGEIPSLKYKDIDWTPTTKEEMYYYQIETLLKLGISQLDCCRYNNDKRAIVDSGTT 307  
OY 249 NLRLPKRYEEAAVYSIKAASTEFKFPDGFMLGDELVCVMQAGTTPWNTEFYISLYLMEGYT 308  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : |  
Db 308 LLRLPQKFVDVAVARASLIPEESGCFMTGSOLACWTSETSPWSFXFKISYLDEDNS 367  
OY 309 NQSFRTITLPOOLRPVEDVATSDDCDKAISISSTGTGMAGVIMGEFVVEDRRARKRI 368  
:::|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |:  
Db 368 SRSPRITILPOLYIQPMMGAGLV-ECYRGISBSTNALVIQATVMEGFIIVIDRQKKRY 426  
OY 369 GFAYASACHVHEFTAAVEGFPVTLDM 395  
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : |  
Db 427 GFAASPCEAIELAGAAVSIEISGFEFTEDV 453

```

RESULT      4
US-09-434-427-2
: Sequence 2, Application US/09434427
: Patent No. 6162630
: GENERAL INFORMATION:
: APPLICANT: POWELL, DAVID J.
: APPLICANT: SOUTHAN, CHRISTOPHER
: APPLICANT: CHAPMAN, CONRAD G.
: APPLICANT: EVANS, JOANNE R.
: TITLE OF INVENTION: ASPI
: FILE REFERENCE: GH-70262-D1
: CURRENT APPLICATION NUMBER: US/09/434,427
: CURRENT FILING DATE: 1999-11-04
: EARLIER APPLICATION NUMBER: US 08/999,723
: EARLIER FILING DATE: 1997-10-06
: EARLIER APPLICATION NUMBER: UK 9626022.9
: EARLIER FILING DATE: 1996-12-14
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 518
: TYPE: PR1T
: ORGANISM: HOMO SAPIENS
: US-09-434-427-2

```

Query Match	52.5%	Score 1132;	DB 4;	Length 518;
Best Local Similarity	54.0%;	Pred. No. 4.1e-113;		
Matches 209; Conservative	65;	Mismatches 109;	Indels 4;	Gaps 2;

[illegible]

Db 368 SRSPFIIITLLPOLYIQPMKGAGLNT-ECYREGISPTNALYIGATVMEGFVIEDRQKRV 426

QY 369 GFAVSACHVHDEFTAAVEGFPVLLDM 395

427 GFAASPCAETLGAANVSFELSGCFSTEDV 453

```

RESULT      5
US-09-717-432-2
: Sequence 2, Application US/09717432
: Patent No. 6291223
: GENERAL INFORMATION:
:   APPLICANT: ZHU, YUAN
:   APPLICANT: LI, XIAOTONG
:   APPLICANT: CHRISTIE, GARY
:   APPLICANT: POWELL, DAVID J.
:   TITLE OF INVENTION: Mouse Aspartic Secretase-1 (maspi)
:   FILE REFERENCE: GP-70663
:   CURRENT APPLICATION NUMBER: US/09/717,432
:   CURRENT FILING DATE: 2000-11-21
:   PRIOR APPLICATION NUMBER: 60/166,974
:   PRIOR FILING DATE: 1999-11-23
:   NUMBER OF SEQ ID NOS: 2
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 2
:   LENGTH: 514
:   TYPE: PRT
: ORGANISM: MUS MUSCULUS
: US-09-717-432-2

```

Query Match	51.8%;	Score 1116.5;	DB 4;	Length 514;
Best Local Similarity	53.7%;	Pred. No. 1.9e-111;		
Matches 209; Conservative	66;	Mismatches 109;	Indels 5;	Gaps 3;

[illegible]

RESULT 6  
US-09-912-484-2  
Sequence 2, Application US/09912484  
Patent No. 6386725  
GENERAL INFORMATION:  
APPLICANT: Christie, Gary  
APPLICANT: Li, Xiaotong  
APPLICANT: Powell, David J.  
APPLICANT: Zhu, Yuan

TITLE OF INVENTION: Mouse Aspartic Secretase-1 (MASP1)  
 FILE REFERENCE: GP-70663-D1  
 CURRENT APPLICATION NUMBER: US/09/912.484  
 CURRENT FILING DATE: 2001-07-25  
 PRIOR APPLICATION NUMBER: 60/166,974  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: 09/717,432  
 PRIOR FILING DATE: 2000-11-21  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: FASTSEQ for Windows Version 3.0  
 SEQ ID NO 2  
 LENGTH: 514  
 TYPE: PRT  
 ORGANISM: MUS MUSCULUS  
 US-09-912-484-2

Query Match 51.8%; Score 1116.5; DB 4; Length 514;  
 Best Local Similarity 53.7%; Pred. No. 1.9e-111;  
 Matches 209; Conservative 66; Mismatches 109; Indels 5; Gaps 3;

8 EPGR-RGSFVEVMDLNRKSGGGYVEMTVGSPPTQTLNLIVDTGSSNFAVGAAPPELHR 66  
 65 EYRATANTLANVADMLQGDSSGRGYLEMLIGTPQKQILVDTGSSNFAVGAAPHSYIDT 124  
 67 YQRQLSTYRDLRKGYVYPTQKMEGLGTDLYSIPHPNVTYRANAIATGSDKFEI 126  
 125 YDSSSSSTYHSKGFVDYKYTGSMTEGVDLVITTRKGFSSFLVNATTFESENFLL 184  
 127 NCSNMEGILGLVAEIAEPDSDLEPFDSLVKQTHVNPILFSLQCGAGFPLNQSSEVLASV 186  
 185 PEIKNGILGLVAYALAKSSLEPFDSLVQAKIPDIFSMQCGAGLPVAGS--GTN 241  
 187 GSGMIIIGIDSLYGLSLMYTPIRREMYEVLIVREINGQDLKMDCKRYNDKSLVDSG 246  
 242 GGSIVIGIGIEPLKYGDIWYTPRIKEMYYQIEILKLEIGOMINLDCREYNADKALVDSG 301  
 247 TTNLRPKVFEAAVKSIAASTEKPPDGFMLGEOLVCMQAGTPMNFPIYSILYMG 306  
 302 TLLRTPQVDFVAVAVARTSLIPEFSDGFMTGAOLACTMSETPMAFPKISITLRDE 361  
 307 VNOSFRITILPQOYLRPVEDVATSDDCYKFAISQSSGTGYWAGVIMGEFYVFDRAK 366  
 362 NMSRFRITILPQYLITQPMGAGFN-ECYRFGISSTNALVIGATVMGEFYVFDRAOR 420  
 367 RIGFAVSACHVDEFTAAVEGFVTLDM 395  
 421 RVGFVAVSPCAIEGTIVSEISGPFSTEDI 449

RESULT 7  
 US-08-208-007A-13  
 Sequence 13, Application US/08208007A  
 Patent No. 5501969  
 GENERAL INFORMATION:  
 APPLICANT: HASTINGS, ET AL.  
 TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CARBILA, BYRNE, BAIN, GILFILLAN,  
 ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 STREET: 6 BECKER FARM ROAD  
 CITY: ROSELAND  
 STATE: NEW JERSEY  
 COUNTRY: USA  
 ZIP: 07068  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH DISKETTE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WORD PERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/208, 007A

FILING DATE: March 8, 1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: NO. 5501969e  
 FILING DATE: NO. 5501969e  
 ATTORNEY/AGENT INFORMATION:  
 NAME: FERRARO, GREGORY D.  
 REGISTRATION NUMBER: 36,134  
 REFERENCE/DOCKET NUMBER: 325800-95  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 396 AMINO ACIDS  
 TYPE: AMINO ACID  
 STRANDEDNESS:  
 TOPOLOGY: LINEAR  
 MOLECULE TYPE: PROTEIN  
 US-08-208-007A-13

Query Match 13.9%; Score 299.5; DB 1; Length 396;  
 Best Local Similarity 25.9%; Pred. No. 1.2e-23;  
 Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

3 DEEPEPGRGSEVMDLNRKSGGGYVEMTVGSPPTQTLNLIVDTGSSNFAVGA---- 58  
 63 DOSAKP-----LIVYLD-----MEYFGTISIGSPQNTVFEDTGSSMLWPSVYCT 110  
 59 AHPFLHRYQRQLSTYRDLRKGYVYPTQKMEGLGTDLYSIPHPNVTYRANAI 118  
 111 SPACKTHSRFQPSQSTSYSOPOQSFSLGYGSLGIIAGQVSV-EGITVVGQDFGESV 169  
 119 TESDFEINGNWEGILGLVAEIAEPDSDLEPFDSLVKQTHVNPILFSLQCGAGFPLN 178  
 170 TEPQTFVD-AEFDLILGIFPSLA--VGGYTPVDNMAQ-----NLVDLPMFSYMSN 222  
 179 QSEVLASVGGSMIIIGIDSLYGLSLMYTPIRREMYEVLIVREINGQDLKMDCKRYN 238  
 223 PE---GGAGSELIIFGYDHSRSGSLNWPVTKAWQIADLNIGVG--TFMFCSE--G 275  
 239 DKSIVDSGTTNLRPKKVFEEAAVKSIAASTEKPPDGFMLGEOLVCMQAGTPMNFPIV 298  
 276 COAIVDTGTSILITGPSDKIKOLQNAIGAP-----VDEYAVE-----CANLVMPD 322  
 299 ISLYLMGEVTNOSFRITILPQOYLRPVEDVATSDDCYKFAISQSSGT----- 346  
 323 VFTTNG-----VPTTSLPTAV--TLIDFVDMQFC-----SSGFGGLDIHPAPG 366  
 347 -TVMGAVIMEGFYVFDRAKRIGFA 371  
 367 LMTLGDVFIROFYVFPDRGNRRVGLA 392

RESULT 8  
 US-09-032-523-9  
 Sequence 9, Application US/09032523  
 Patent No. 6232454  
 GENERAL INFORMATION:  
 APPLICANT: Bandman, Olga  
 APPLICANT: Hillman, Jennifer L.  
 APPLICANT: Corley, Neil C.  
 APPLICANT: Guegler, Karl  
 APPLICANT: Baugh, Mariah  
 TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA











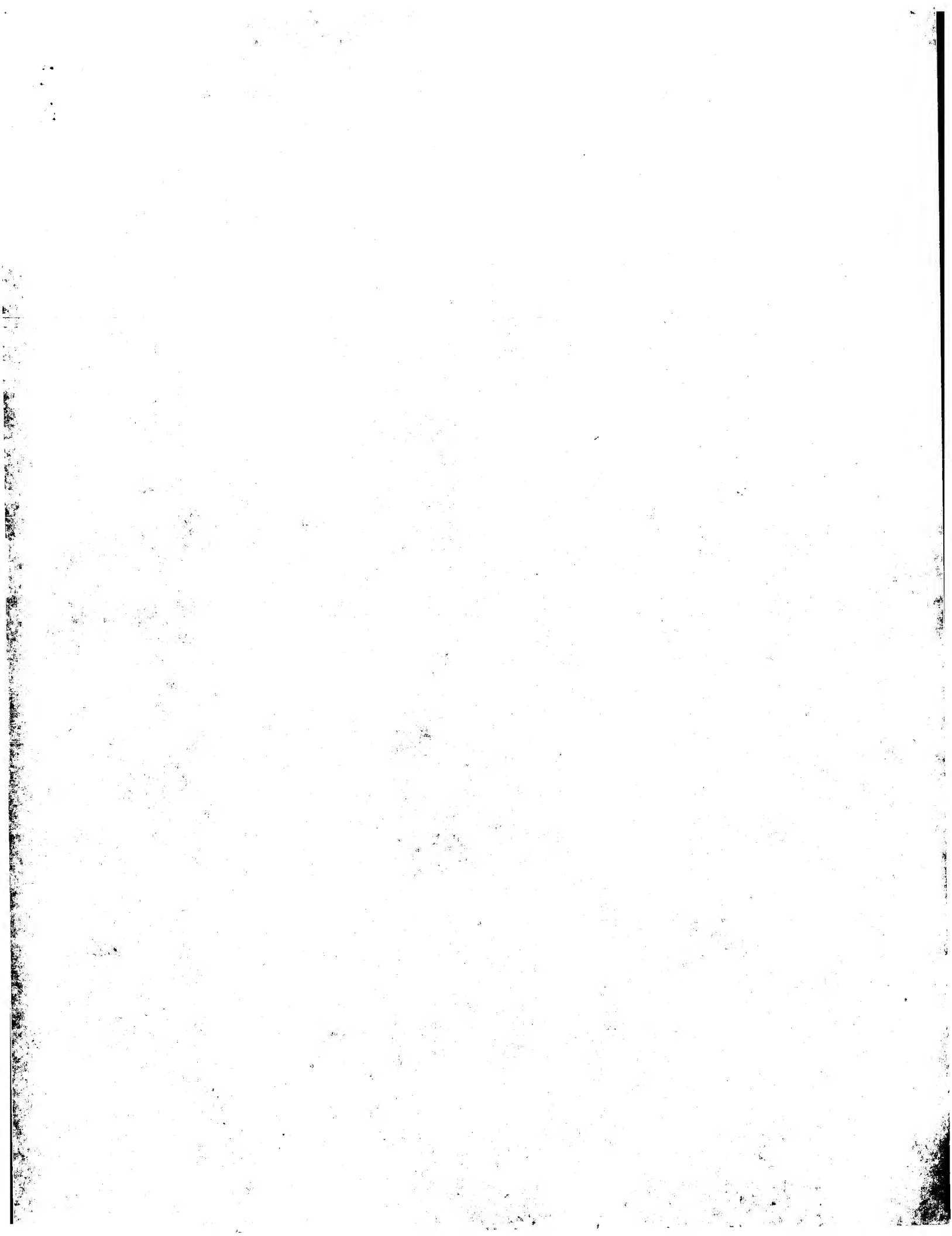
Wed Oct 30 14:11:51 2002

us-09-724-571-58.ra1

Page 9

DB 357 DYTEVSGSCISAFPMDFP-EPISPLAIGDSFLRKYYVYDLGKDAVGLAKS 409

Search completed: August 7, 2002, 09:14:49  
Job time: 74 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:14:17 ; Search time 32.31 Seconds

(without alignments)  
1210,410 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 2156

Sequence: 1 ETDEPEPEGRGSGFYEMVD.....GPFVTLDMEDCGYNIPQTE 407

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

PIR\_71: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2156	100.0	501	2 A59090	aspartic proteinase
2	308	14.3	384	2 JC7574	pepsinogen A - Afr
3	308	14.3	385	2 JC7575	pepsinogen A - bul
4	306	14.2	387	2 B38302	pepsinogen C - Afr
5	303	14.0	388	1 JC7573	pepsinogen C - Afr
6	302.5	14.0	388	1 S19682	pepsin A (EC 3.4.2
7	301	14.0	382	1 PECH	pepsin A (EC 3.4.2
8	299.5	13.9	396	2 A34401	cathepsin E (EC 3.
9	298.5	13.8	383	2 A14443	pepsin (EC 3.4.23.
10	298.5	13.8	384	2 A39314	gastricsin (EC 3.4
11	296	13.7	387	2 C38302	pepsin (EC 3.4.23.
12	296	13.7	391	2 A43356	cathepsin E (EC 3.
13	295.5	13.7	412	1 KHHUD	cathepsin D (EC 3.
14	295	13.7	387	2 D38302	pepsin A (EC 3.4.2
15	289.5	13.4	388	1 S19684	pepsin A (EC 3.4.2
16	287.5	13.3	444	2 T24204	hypothetical prote
17	287	13.3	407	1 KHRD	cathepsin D (EC 3.
18	286	13.3	387	2 E38302	pepsin (EC 3.4.23.
19	285.5	13.2	398	2 S66465	cathepsin E (EC 3.
20	285	13.2	398	2 S66465	cathepsin E (EC 3.
21	283.5	13.1	388	1 S11185	cathepsin D (EC 3.
22	282.5	13.1	388	1 PEMOAR	pepsin A (EC 3.4.2
23	281.5	13.1	381	1 CSMHR	pepsin A (EC 3.4.2
24	281.5	13.1	388	1 PEMOAJ	pepsin A (EC 3.4.2
25	281.5	13.1	388	2 A30142	pepsin A (EC 3.4.2
26	279.5	13.0	388	2 B30142	pepsin A (EC 3.4.2
27	279.5	13.0	410	1 KHMSP	cathepsin D (EC 3.
28	278.5	12.9	386	1 PEPG	pepsin A (EC 3.4.2
29	278	12.9	387	2 JC7245	pepsinogen A - com

30	277.5	12.9	380	2 I47176	chymosin (EC 3.4.2
31	277.5	12.9	396	2 S36865	cathepsin E (EC 3.
32	276	12.8	389	2 JE0371	pepsin C (EC 3.4.2
33	273.5	12.7	381	1 CMBO	chymosin (EC 3.4.2
34	270.5	12.5	377	1 PEMOAJ	gastricsin (EC 3.4
35	270.5	12.5	389	2 A38302	pepsin (EC 3.4.23.
36	270	12.5	376	2 I45856	aspartic proteinase
37	268.5	12.5	344	1 KHPGD	cathepsin D (EC 3.
38	267.5	12.4	381	2 JC7247	prochymosin - comm
39	266	12.3	380	2 S03433	saccharopepsin (EC
40	266	12.3	405	2 A25379	aspartic proteinase
41	264	12.2	396	2 T47207	pepsinogen C - com
42	263.5	12.2	388	2 JC7246	gastricsin (EC 3.4
43	262.5	12.2	394	2 B43356	aspartic proteinase
44	261.5	12.1	387	2 A45117	gastricsin (EC 3.4
45	261.5	12.1	388	2 A29937	gastricsin (EC 3.4

#### ALIGNMENTS

##### RESULT 1

A59090

aspartic proteinase (EC 3.4.23.-) BACE precursor - human

N/Alternate names: beta-secretase; beta-site APP cleaving enzyme

C:Species: Homo sapiens (man)

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-May-2000

C/Accession: A59090

R/Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendar, E.A.; Denis, P.; Teplio

M.A.; Blere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro

Science 286, 735-741, 1999

A>Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran

A/Reference number: A59090; MUD:20002972

A/Note: Submitted to Genbank, September 1999

A/Accession: A59090

A>Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-501 <VAS>

A/Cross-references: GB:AF190725; NID:96118538; PIDN:AAF04142.1; PID:96118539

C:Genetics:

A:Gene: BACE

C/Superfamily: beta-secretase

C/Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-45/Domain: propeptide #status predicted <PRO>

F:46-501/Product: acid proteinase BACE #status predicted <MAT>

F:461-477/Domain: transmembrane #status predicted <TRN>

F:93-289/Active site: Asp #status predicted

F:153,172,223,354/Binding site: carbohydrate (asn) (covalent) #status predicted

F:330-380/Disulfide bonds: #status predicted

Query Match 100.0%; Score 2156; DB 2; Length 501;

Best Local Similarity 100.0%; Pred. No. 1,6e-173;

Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ETDEPEPEGRGSGFYEMVDNLKSGGYVEMTQSPQNTIIIVDGSNFAVGAAP	60
DB	46	ETDEPEPEGRGSGFYEMVDNLKSGGYVEMTQSPQNTIIIVDGSNFAVGAAP	105
QY	61	HPPLHRYQOLSTYRDLKRGYVYPTQGWGELGTDLVSIHPGPNVTRANIATTE	120
DB	106	HPPLHRYQOLSTYRDLKRGYVYPTQGWGELGTDLVSIHPGPNVTRANIATTE	165
QY	121	SDRFINGNMGITGLAAVEIRAPDLSLEPPDSLVKQTHVNFSLDLCAGAPPLNS	180
DB	166	SDRFINGNMGITGLAAVEIRAPDLSLEPPDSLVKQTHVNFSLDLCAGAPPLNS	225
QY	181	EVLAVGGSMITIGIDHSYTGSLMTPTIRREMYEVIIVREINGODLKMCKEYNDX	240
DB	226	EVLAVGGSMITIGIDHSYTGSLMTPTIRREMYEVIIVREINGODLKMCKEYNDX	285
QY	241	SLVDSGTNLRPLKVFEEAFAVSKRAASTEFKPDGFWLGEOLVCMQAGTTPWNIFFVIS	300

```

|||||
Db 286 SIYDSGTNLRKRYEAVKSIKAASSTKRPDGFGLGEOLVCMQAGTTPNNIPVIS 345
QY 301 LYLMGEVYTNOSFRITLIPQOYLRPVEDVATSODCYKFAISOSTGTWGAIVMEGFYV 360
Db 346 LYLMGEVYTNOSFRITLIPQOYLRPVEDVATSODCYKFAISOSTGTWGAIVMEGFYV 405
QY 361 FDRARRKRGFAVSACHVDEFRFAAAGSPVTLDMEDCGINIPQTD 407
Db 406 FDRARRKRGFAVSACHVDEFRFAAAGSPVTLDMEDCGINIPQTD 452

RESULT 2
JC7574
pepsinogen A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: J07574; PC7119
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: J07573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: J07574
A:Molecule type: mRNA
A:Residues: 1-384 <IKU>
A:Cross-references: DDBJ:AB045380
A:Accession: PC7119
A:Molecule type: protein
A:Residues: 16-35;57-76 <IK2>
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: PGA
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 14.3%; Score 308; DB 2; Length 384;
Best Local Similarity 27.5%; Pred. No. 4.8e-18;
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;

QY 30 YVEMTVGSPPTNLINLVDTGSSNFAGAPHPFL-----HRYQRLSTYRDLKRG 82
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 YGTGISGTPQEFYVIFDTGSANLW---PSYSSQACSNHNRFPQOSTFOATNTP 128
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 83 VYVYTGKWEGLGDLVSIPIGPNVTVRANIAITTESK-FFINGSNMEGLGLAYAE 141
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 129 VSIQYTGSMGSLGIDYDLOV---GNIQISNOMGSLSESPGSLYSPDGILGLAFPS 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 142 IARDDSLPEFDSLVRKQTHVP-NLFSIQCGAGFPINOSEVLASVGSMLIGIDHSLY 200
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 IA--SSQATPFEDMMWSQGLIPQNLFSYILSSDG-----QTSYLVFGVDNSY 233
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 201 TGSIMVTPIRREYVEVIVVEINGDPL-KMDCKEYNKDSIVDSGTNLRPKKVF 258
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 SGSLMWVLTETWQITLDVSVISGVIAOSOSC-----QALVDGTSLMTGSPPT- 286
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 259 AAVKSIRKASSTKFPDGFMLGEOLV-CMQAGTTPWNIFVYISLYLMGEVYTNOSFRITIL 317
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 287 ANIQWYIGASODSN-----GQYVINCNNISNMPTVF-----TIN 321
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 318 POQY-LRPVEDVATSODDCYK-FAISOSTGT---VMGAVIMEGFYVFPDRARRKRGFA 371
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 322 GVQYPLSPFAVYKQNOQCGSSGFQAMNLPITNSGDLMIIGDVFIRQYFTVFPDRANNYVIA 381
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
JC7575
pepsinogen A - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: J07575
R:Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.

```

```

J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogen
A:Reference number: J07573; MUID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: J07575
A:Molecule type: mRNA
A:Residues: 1-385 <IKU>
A:Cross-references: DDBJ:AB045376
A:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: PGA
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

Query Match 14.3%; Score 308; DB 2; Length 385;
Best Local Similarity 27.8%; Pred. No. 4.8e-18;
Matches 99; Conservative 64; Mismatches 133; Indels 60; Gaps 15;

QY 30 YVEMTVGSPPTNLINLVDTGSSNFAGV---AAPHPILHRYQRLSTYRDLKRGYV 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 73 YGTGISGTPQEFYVIFDTGSSNLMWPSYSSQACNTNHHMFNPQOSTFOATNTPVSI 132
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 86 PYTGKWEGLGDLVSIPIGPNVTVRANIAITTESDK-FFINGSNMEGLGLAYAEIAR 144
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 133 QYGTGSMGSLGIDYDLOV---GNIQINQIFGLSOSPEPSFLYSPDGILGLAFPSLA 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 145 PDDSLPEFDSLVRKQTHVP-NLFSIQCGAGFPINOSEVLASVGSMLIGIDHSLYTGS 203
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 189 -SSQATPFEDMMWSQGLIPQDLFSYILSSOG---OS-----GSFVLEFGVDNSYTN 237
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 204 LMYPIRREYVEVIVVEINGDQDKM--DCKEYNKDSIVDSGTNLRPKKVEAAV 261
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 LNWVLTETWQITVVISIGQVIAOSGC-----SAIVDGTSLMGF---STPI 287
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 262 KSIRKASSTKFPDGFMLGEOLVCMQAGTTPWNIFVYISLYLMGEVYTNOSFRITILPQY 321
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 288 ANIQYIIGANDSNQYV---INCNNISNMPTVF-----TINCQY 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 322 LRPVED-VATSODDC---YKFAISOSTGT---VMGAVIMEGFYVFPDRARRKRGFA 371
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 327 PLPASAVYKQSOQCTSGFQAMNLPITSSGDLMIIGDVFIRYVYFPDRANNYVAMA 382
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
B38302
pepsin (PC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Kolwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
A:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 14.2%; Score 306; DB 2; Length 387;
Best Local Similarity 27.1%; Pred. No. 7.2e-18;
Matches 96; Conservative 67; Mismatches 131; Indels 66; Gaps 15;

QY 30 YVEMTVGSPPTNLINLVDTGSSNFAGV---AAPHPILHRYQRLSTYRDLKRGYV 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 YGTGISGTPQEFYVIFDTGSSNLMWPSYSSQACSLACFLKRNPPDSSFTFOATSTLSL 134
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 86 PYTGKWEGLGDLVSIPIGPNVTVRANIAITTESD---KFFINGSNMEGLGLAYAEI 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 135 TYGSGMTGILGDTYKV---GNIEDTNOIGLSKTEPGITFLV--APFDGILGLAVPSI 189
Oy 143 ARPDDSLPEFFDSLVKQTHV--PNLFSIQLCGAGPFLNSELVAVSGMIIIGIDHSIYT 201
Db 190 SASDAT--PVEDNMNNEGLVEDLFSVYLSNG-----EKSGVMWGGIDSSYTT 237
Oy 202 GSLMTPTLRREMYEVIIIVRVEINGODIKM--DKEYNYDKSIYDSGTTNLRKKVFEA 259
Db 238 GSLMWPVSHGEGYWOITMDSITINGETIACADSC-----QAVVDGTSLAGPTSAISK 291
Oy 260 AVKSIKAASTKEKFPDGFWMLEOLV--CMQAGTTPMNIFFPVISLYLMEGVINQSFRTILP 318
Db 292 IQSYIGASKNL-----LGEITISCAIDSLPDIVF-----TINN 325
Oy 319 QOYLRPEVD--VATSODDC---YKFAISQSTGT--VMGAVIMEGFYVFPDRARRIGFAV 372
Db 326 VOYPLPASAVYILKEDDCLSGFDCGMNDTYSGLMIIIGDVFIRQYTFVFRANNQVGLAA 385
Oy 373 SA 374
Db 386 AA 387

```

```

RESULT 5
JC7573
pepsinogen C - African clawed frog
N:Alternate names: progastricin
C:Species: Xenopus laevis (African clawed frog)
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: J07573; PC7118
R:Kuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasa, S.
J. Biochem. 129, 147-153, 2001
A:Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens
A:Reference number: J07573; M0ID:21064922; PMID:11134969
A:Contents: Stomach
A:Accession: J07573
A:Molecule type: mRNA
A:Residues: 1-388 <IKU>
A:Cross-references: DDBJ:AB045379
A:Accession: PC7118
A:Molecule type: protein
A:Residues: 17-68 <IK2>
C:Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like
C:Genetics:
A:Gene: Pgc
C:Superfamily: pepsin
C:Keywords: stomach; zymogen

```

```

Query Match 14.1%; Score 303; DB 2; Length 383;
Best Local Similarity 27.4%; Pred. No. 1.3e-17;
Matches 104; Conservative 57; Mismatches 122; Indels 96; Gaps 17;

Oy 30 YVEMTVGSPPTLNLIVDTGSSNFAVGA-----APHPLHRYORULSTYRDLRK 81
Db 67 YVGEISIGRPONFLVLDGTSSNLMVASTYQSOQACTNHP-----FNPSSQSTYSSNQ 122
Oy 82 GYVYVYTOGKMGELGTDIVSIPGPNVTYRANIAITAESDKFLNGSN-----WEGTLG 136
Db 123 QPSLQGTGSLGILIGDTVTI---QNAVISQGEFLSETEP---GTNFFYAFQFDGILG 175
Oy 137 LAVAEIARPDLSLEFFDSLVKQTHV--PNLFSIQLCGAGPFLNSELVAVSGSMIIIGID 196
Db 176 LAVPSIA--VGSATTVMGMMQO---NLINQPI--FEGYLSGGS--SQNGEVAFAFGVD 225
Oy 197 HSLYGSLSMTPIRREMYEVIIIVRVEINGOD---LKMDCKEYNYDKSIYDSGTTNLRP 253
Db 226 QNYIYGQITWIPVSEIYTWQIGQFSTNGQATGMCSCGC-----QAIYDTGTSLLTAP 279
Oy 254 KKEVAAYKSIKAASTKEKFPDGFWMLEOLVCMQAGTTPMNI--FPVISLYLMG----- 305
Db 280 QSVFSLISQISGAODQN-----GQYVYVSGS-----NIQNLPTTISISVSFPLP 325

```

```

Oy 306 ---EVTNQS-----FRITLPEQYLRPEVDVATSQDDCYKFAISQSTGTVMGAVIME 355
Db 326 PSAYVLQSSSGYCTIGMPTFLPSQNGPL-----WILGDVFLR 364
Oy 356 GFYVFPDRARRIGFAVSA 374
Db 365 EYISYVIDLGNNOVGAFATAA 383

```

```

RESULT 6
pepsin A (EC 3.4.23.1) 4 precursor - Japanese macaque
N:Alternate names: pepsinogen A isozyme 4
C:Species: Macaca fuscata (Japanese macaque)
C>Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
C:Accession: S19682; S16065
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
Eur. J. Biochem. 202, 205-215, 1991
A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and st
A:Reference number: S19681; M0ID:92037645
A:Accession: S19682
A:Molecule type: mRNA
A:Residues: 1-388 <KAG>
A:Cross-references: EMBL:X59753; NID:938070; PIDN:CAA2425.1; PID:938071
A:Note: Parts of sequence, including amino ends of pepsinogen and activation intermed
C:Comment: This is a minor component of pepsin at all post-partum stages.
C:Superfamily: Although two-step activation is observed, activation is predominantly a o
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein di
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-388/Product: pepsinogen A 4 #status experimental <APY>
F:63-388/Product: activation peptide #status experimental <APY>
F:38-39/Cleavage site: pepsin A 4 #status experimental <EN2>
F:62-63/Cleavage site: Leu-Iys (pepsin) #status experimental
F:94,277/Active site: Asp #status predicted
F:107-112,268-272,311-344/Disulfide bonds: #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

```

```

Query Match 14.0%; Score 302.5; DB 1; Length 388;
Best Local Similarity 27.6%; Pred. No. 1.4e-17;
Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;

Oy 3 DEEPEPGRSGFEVEMDNLRKSGQGYVEMTVGSPPTLNLIVDTGSSNFAVGAAPHP 62
Db 64 DEQPLE-----NYLDV-----EYFGTIGTRPAQNFYVFDGSSNLWV---PSV 105
Oy 63 FL-----HRYORULSTYRDLRKGVYVYTOGKMGELGTDIVSIPGPNVTYRANI 115
Db 106 YCYSIACMDHNLFPQDSSYTRATSKVSTLYTGSGMTGILGDTYKV---GGISDTNQT 162
Oy 116 AAIRESK--FPINSNNEGILGAYAEIARPDLSLEFFDSLVKQTHV--PNLFSIQLCGA 173
Db 163 FGLSETPGFFLFPAPDGLGLAYPSIS--SSGATPVFDNINNRIVSODLFSVYLSAD 220
Oy 174 GFPLNSELVAVSGSMIIIGIDHSIYGSLSMTPIRREMYEVIIIVRVEINGODL--KM 231
Db 221 ----DQS-----GSVYIFGIDSSYTGSLMWPVSVESGYWQISVDSITMNGKTATCAK 270
Oy 232 DCKEYNTDKSIYDSGTTNLRPKVFEAAVKSIRKAASTKEKFPDGFWMLEOLV--CMQACT 290
Db 271 GC-----QAIYDTGTSLLGTPSPIANIQSDIGASNSD-----GEMVYSGAIS 316
Oy 291 TPWNIFPVISLYLMGEVTNQSFRITLIPQY--LRPEVDVATSQDDCYK-----FAISQS 344
Db 317 LPIDIV-----TNGVQYPLPASAYILQSGCTSGFQGMVPTESG 358
Oy 345 TGVVAGAVIMEGFYVFPDRARRIGFA 371
Db 359 ELWILGDVFIRQYTFVFRANNQVGLA 385

```



```

Db 223 PE--GGAGSELIFGVDHSHFSGSLNMVPTKQAYMQIALDNIQVG--TVMFCSE--G 275
QY 239 DKSIVDSGTTNLR.LPKKVEAAVRSIKASSTKPKPGFWMGEGLVCMQAGTTWNNIPV 298
Db 276 CQALVDTCTSLITRPSDKIKQLQNAIGAP-----VGEYAVE-----CANLNMVDP 322
QY 299 ISLILMEGVNTQSFRTITLLPOQLRPVEDVATSODDCYKFAISQSTNG----- 346
Db 323 VTFTING-----VPYLSLPTAY--TLLDPYDGMQFC-----SSGFGGLDHPAPGR 366
QY 347 -TVMGAVIMEGEFVYVPPARRKRIGFA 371
Db 367 LMITGDVFIRKOFISVFDGRNNRVGLA 392

RESULT 9
pepsin (EC 3.4.23.-) precursor, embryonic - chicken
C:Species: Gallus gallus (chicken)
C>Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 21-Jul-2000
C:Accession: A41443
R:Hayashi, K.; Agata, K.; Mochi, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.
J. Biochem. 103, 290-296, 1988
A:Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pef
A:Reference number: A41443; M01D:88227903
A:Accession: A41443
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-383 <HAY>
A:Cross-references: GB:D00215; NID:q2760810; PIDN:BA00153.1; PID:q222853
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 13.8%; Score 298.5; DB 2; Length 383;
Best Local Similarity 25.2%; Pred. No. 36-17;
Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY 30 YYEMVTVGSPQTLNLTVDGSSNFAVGA---APHPFLHRYVQRLSSTYRLRKGVV 85
Db 76 YGTGTISGTPPOFTVFPDTSNLMVPSVCTSPACQSHQMFNPQSSTYSTGQNLSTI 135
QY 86 PYTGKMEGELGTDLYSIPHPNVTVRANIAITTESDKFINGSNMGEGLGLAYAEIARP 145
Db 136 HYGTCGMEGVGCDYTVASLMDTNQLEGLST-SEPGQFVY-VKFDGILGGLPSLA- 192
QY 146 DDLSEPFDSLAKQTHV-DNLFSLQCGAGFPLNQSEVLASVGSMTIGDHSIYTGSL 204
Db 193 -DGITPFDNMVNESLLEQNLFSVLS-----REPMGSMVRFQIDESTFTGSI 240
QY 205 WYPIREKRYEVIIVRVEINQDL--KMDCKEYNDKSIYDSGTTNLR.LPKKVEAAV 262
Db 241 NMLPVSYGQYQWQISMDSTIVNKEIACSSGC-----QALIDTGLSLVAGPASDINDIQS 294
QY 263 SIKASSTKPFDFGFWLGEQLVCMQAGTTPWNIIPVLS-----YLMGEVNTQSFRTITLP 318
Db 295 AVGANQNT-----YGEYSV-----NCSHILAPDVVYFVIGG----- 326
QY 319 QOYLKRVLEVA---TSQDDCYKFAISQSTNGTVMGAVIMEGEFVYVPPARRKRIGFA 371
Db 327 -OY--VPALAVTEONGQGTCMSSFQNSSADLMLIGDVFIRYVSIIDPRANNRVGLA 380

RESULT 10
gastriclin (EC 3.4.23.3) precursor - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C>Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 22-Jun-1999
C:Accession: A39314
R:Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageyama, T.
J. Biol. Chem. 266, 22436-22443, 1991
A:Title: Purification, characterization, and amino acid sequences of pepsinogens and pep

```

```

A:Reference number: A39314; M01D:92042186
A:Accession: A39314
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-384 <YAK>
A:Cross-references: GB:M73750; NID:q213687; PIDN:AAA49530.1; PID:q213688
C:Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 13.8%; Score 298.5; DB 2; Length 384;
Best Local Similarity 25.5%; Pred. No. 36-17;
Matches 100; Conservative 61; Mismatches 142; Indels 89; Gaps 15;

QY 14 SFYEMVDNLKSGSGGYVEMTVGSPQTLNLTVDGSSNFAVGAAPHPFL-----HR 66
Db 51 NFATAFEPLANYMDMSYGEISIGTPPOFTVLEFDGSSNLMV---PSYQCSQACTNHP 107
QY 67 YVQRLSSTYRLRKGVVYVPTQKMEGELGTDLYSIPHPNVTVRANIA-----AI 118
Db 108 QRPSSQSSSYSSNQOQFSIQYGTGSLTGILGYDVOI-----QNALISQGEFGLSV 158
QY 119 TESDKFINGSNMGEGLGLAYAEIARPDLSLPEFDSLAKQTHVFN-LFSQLCGAGFPL 177
Db 159 TEPGTFVYV-AQFDGILGLAYPSIA--EGATTVNQMGIQNLINQPLFAFLYSQGNQSQ 215
QY 178 NOSEVLASVGSMTIGDHSIYTGSLWYTPIRREMYEVIIVRVEINQD---LKMCK 234
Db 216 N-----GGEVAVGAGVDQNNYSQIYWTPTSETYMQIGFQSVNGQATGMSQSC- 266
QY 235 EYNDKSIYDSGTTNLR.LPKKVEAAVRSIKASSTKPKPGFWMGEGLV-CMQAGTTPW 293
Db 267 -----QGIYDTCTSLITRPSQSVFSSLMQSIGAQDQD-----GOYAVSCSNQSLPT 313
QY 294 NIFPVI-----SLYLMGEVNTQ--SFRITLIPQQLRPVEDVATSODDCYKFAISQ 342
Db 314 ISFTISGVSPFLPPSAVYVLDQNSGYCTIGIMPTPLPQNGQPL----- 356
QY 343 SSTGTVMGAVIMEGEFVYVPPARRKRIGFAVSA 374
Db 357 ----WILDGDFLROYSYVDLGNNOVGFMAAA 384

RESULT 11
pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997
C:Accession: C38302
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; M01D:91009127
A:Accession: C38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 13.7%; Score 296; DB 2; Length 387;
Best Local Similarity 26.9%; Pred. No. 56-17;
Matches 97; Conservative 63; Mismatches 135; Indels 66; Gaps 13;

QY 30 YYEMVTVGSPQTLNLTVDGSSNFAVGAAPHPF-----LHRYVQRLSSTYRLRK 82
Db 75 YFGTISGTPPOFTVLEFDGSSNLMV---PSYQCSLACALHKRNPEDSSYTGCTSET 131
QY 83 VVYVYTGKMEGELGTDLYSIPHPNVTVRANIAITTESKPFINGSNMGEGLGLAYAEI 142
Db 132 LSLITYGSMTGILGIDTVKAVGSIEDTNOITFGLSKTEPSLTLFL--APPDGILGLAYPSI 189

```

```

OY 143 ARPDSLEPPFDSLVQOTIV-PIEISLQCGAGPFLNQSEVLASVGGMIIIGIDHSLYT 201
Db 190 SSSPAT--PFDNMNMEGLASODLFSVLYSSDD-----EKGSLVMGIGDSSYTT 237
OY 202 GSLMYTPFIREWYEVYIIVREINGODLKM--DCKEYNDKSIYDSGTTNLRPKKVEEA 239
Db 238 GSLNMWVVEYEGVWQJTMDSVSIINGETIACADSC-----QAIYDTGTSLLTGP----TS 287
OY 260 AVMSIKRASTEKFPDGFMTLGBDLV-CWQAGTTPWNIFPVISLYLMEGVTSQSFRTILP 318
Db 288 AISNIQSYIGASK---NLGENVICSALDILPDLVF-----TLNG 325
OY 319 QOYLRPVEDVATSOODCKYFAISQSSTG---VMGAVIMEGVYVVDARRKIRGAV 372
Db 326 IQYPLPASVATLKEDDCTSGLEGMMWVDYTGELWILGDFVTRQYFTVFDRRANQLGLAA 385
OY 373 S 373
Db 386 A 386

```

RESULT 12  
A43356  
cathpsin E (EC 3.4.23.34) precursor - guinea pig  
N:Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase  
C:Species: *Cavia porcellus* (guinea pig)  
C:Date: 31-Dec-1999 #sequence\_revision 31-Dec-1993 #text\_change 22-Jun-1999  
C:Accession: A43356  
R:Kaeyama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koizumi, O.; Tanji, M  
J. Biol. Chem. 267, 14550-14559, 1992  
A:Title: Gastric procathepsin E and progastricsin from guinea pig. Purification, molecular  
A:Reference number: A43356; MUID:9235614  
A:Accession: A43356  
A:Molecule type: mRNA  
A:Residues: 1-391 <KAG>  
A:Cross-references: GB:M86653; NID:9191294; PIDD:AAA37052.1; PID:9191295  
A:Note: sequence extracted from NCBI backbone (NCBIN:110763, NCBIF:110769)  
C:Keywords: aspartic proteinase; hydrolase; membrane protein

Query Match	13.7%;	Score 296;	DB 2;	Length 391;
Best Local Similarity	26.98;	Pred. No. 5.1e-17;		
Matches 98; Conservative	64;	Mismatches 130;	Indels 72;	Gaps 16

RESULT 13

KHND

cathepsin D (PC.3.4.23.5) precursor [validated] - human

N:Alternate names: preprocathepsin D

C:Species: Homo sapiens (man)

C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #extl\_change 15-Sep-2000

C:Accession: A25771; S30749; PC2066; I5236; I57716

R:Faust, P.L.; Kornfeld, S.; Chirgwin, J.M.

Proc. Natl. Acad. Sci. U.S.A. 82, 4910-4914, 1985

A:Title: Cloning and sequence analysis of cDNA for human cathepsin D.

A:Reference number: A25771; MUID:85270436

A:Accession: A25771

A:Molecule type: mRNA

A:Residues: 1-412 <FNU>

A:Cross-references: EMBL:M11233; NID:g181179; PIDN:AAB59529.1; PID:g181180

R:Westley, B.R.; May, F.E.B.

Nucleic Acids Res. 15, 3773-3786, 1987

A:Title: Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive human br

A:Reference number: S30749; MUID:87231068

A:Accession: S30749

A:Molecule type: mRNA

A:Residues: 1-412 <MBE>

A:Cross-references: GB:J12980; NID:g291930; PIDN:AAAL634.1; PID:g455429

R:May, F.E.B.; Smith, D.J.; Westley, B.R.

Gene 134, 277-282, 1993

A:Title: The human cathepsin D-encoding gene is transcribed from an estrogen-regulate

A:Reference number: PC2066; MUID:94085791

A:Accession: PC2066

A:Molecule type: DNA

A:Residues: 1-23 <MA>

A:Cross-references: GB:J12980; NID:g291930; PIDN:AAAL634.1; PID:g455429

A:Experimental source: MCF-7 cell

R:Cavailles, V.; Augereau, P.; Rochefort, H.

Proc. Natl. Acad. Sci. U.S.A. 90, 203-207, 1993

A:Title: Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate

A:Reference number: I59236; MUID:93126342

A:Accession: I59236

A>Status: translation not shown; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-22 <DAV1>

A:Cross-references: GB:S52557; NID:g263124; PIDN:AAD1368.1; PID:g4261568

R:Augereau, P.; Mitalles, F.; Cavailles, V.; Gaudeliet, C.; Parker, M.; Rochefort, H.

Mol. Endocrinol. 8, 693-703, 1994

A:Title: Characterization of the proximal estrogen-responsive element of human cathep

A:Reference number: I57716; MUID:95021301

A:Accession: I57716

A>Status: translation not shown; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-22 <CAV2>

A:Cross-references: GB:S74689; NID:g786350; PIDN:AAD14156.1; PID:g4261856

R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.

Submitted to the Brookhaven Protein Data Bank, April 1993

A:Reference number: A51839; PDB:1LYA

A:Contents: annotation: X-ray crystallography, 2.5 angstroms, residues 65-161,170-241

R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.

Submitted to the Brookhaven Protein Data Bank, April 1993

A:Reference number: A51840; PDB:1LYB

A:Contents: annotation: X-ray crystallography, 2.5 angstroms, with inhibitor residues

R:Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Hosur, M.V.; Sowder, II, R.C.; Cachau, R.E.;

Proc. Natl. Acad. Sci. U.S.A. 90, 6796-6800, 1993

A:Title: Crystal structures of native and inhibited forms of human cathepsin D: impli

A:Contents: annotation: X-ray crystallography, 2.5, angstroms

C:Comment: Cathepsin D is a ubiquitous lysosomal proteolase.

C:Comment: In addition to the propeptide, residues 163-168 and 411-412 are proteolytic

C:Comment: The carbohydrate bound to 134-Asn contains a mannose-6-phosphate that is b

C:Gene: GDB:CTSDD

A:Cross-references: GDB:120512; OMIM:116840

A:Map position: 11p15.5-11p15.5

C:Function:





QY 177 LNOSEVLASVGSMTIGIDHSIYTGSLWYTPIRREMYEVIIVREINGODLKMDCREY 236  
Db 221 -DOS-----GSVYIFGGIDSSYTGSLMWVPYSVEGYWQISYDSITMNGEAI--CAE- 270  
QY 237 NYDKSIYDSGFTNRLPKVFEAAVKSIRKASSTEKFPDGFWLGEOLY-CWQAGTTPWNI 295  
Db 271 -GGOAIYDTSTSLTGFTSPIANIOSDIDGASENSD-----GEMVYSCSAISLDPDIV 321  
QY 296 FPVYSIYLMEGVNOSFRITILPOQYLRPEVDVATSDDCYK-----FAISQSGTGTVMG 350  
Db 322 F-----TINGIQYVPPSAI-----ILQSGSCISGFGQMDVPTESGELMILG 364  
QY 351 AVIMEGFYVVDRAKRRIQFA 371  
Db 365 DVFTROYFTVFDRAANNOVGILA 385

Search completed: August 7, 2002, 09:14:18  
Job time: 42 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:15:11 ; Search time 17.62 Seconds

(without alignments)  
894.373 Million cell updates/sec

Title: US-09-724-571-58

Sequence: 1 ETDEPPEPPGRGSGFVEMVD.....GPFTLMDGCGYNIPQTD 407

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2156	100.0	501	1	BACE_HUMAN
2	2134	99.0	501	1	BACE_RAT
3	2132	98.9	501	1	BACE_MOUSE
4	1132	52.5	518	1	BACE_HUMAN
5	327	15.2	324	1	PEP1_GADMO
6	314.5	14.6	390	1	CARD_BOVIN
7	306	14.2	387	1	PEP1_RABIT
8	302.5	14.0	388	1	PEP4_MACFU
9	302	14.0	367	1	PEP4_CHICK
10	299.5	13.9	396	1	CATE_HUMAN
11	298.5	13.8	383	1	PEP2_CHICK
12	296	13.7	391	1	CATE_RABIT
13	295	13.7	412	1	CATE_HUMAN
14	295.5	13.7	412	1	CATE_HUMAN
15	295	13.7	412	1	CATE_HUMAN
16	289.5	13.4	388	1	PEP2_MACFU
17	287	13.3	407	1	CATE_RAT
18	286	13.3	387	1	PEP3_RABIT
19	285.5	13.2	398	1	CATE_RAT
20	285	13.2	398	1	CATE_RAT
21	283.5	13.1	388	1	PEP4_MACFU
22	282.5	13.1	388	1	PEP4_HUMAN
23	281.5	13.1	381	1	CHYM_SHEEP
24	281.5	13.1	381	1	PEP1_MACFU
25	279.5	13.0	410	1	CARD_MOUSE
26	278.5	12.9	386	1	PEP4_PIG
27	278	12.9	387	1	PEP4_CALJA
28	277.5	12.9	396	1	CATE_RABIT
29	277.5	12.9	397	1	CATE_MOUSE
30	276.5	12.8	419	1	CARD_BOVIN
31	273.5	12.7	388	1	PEP2_RABIT
32	273	12.7	388	1	PEP2_RABIT
33	270.5	12.5	377	1	PEP4_MACFU

34	270	12.5	376	1	PAC2_BOVIN	Q28057 bos taurus
35	267.5	12.4	381	1	CHYM_CALJA	Q9n2d2 callithrix
36	267	12.4	365	1	CARD_SHEEP	Q9ms88 ovis aries
37	266	12.3	405	1	CARD_YEAST	P07267 saccharomyc
38	264	12.2	396	1	CARD_NEUCR	Q01294 neurospora
39	263.5	12.2	388	1	PEP3_CALJA	Q9n2d3 callithrix
40	262.5	12.2	394	1	PEP3_CAVPO	Q64411 cavia porce
41	261.5	12.1	387	1	ASBP_AEDAE	Q03168 aedes aegyp
42	261.5	12.1	388	1	PEP3_HUMAN	P20142 homo sapien
43	261	12.1	388	1	PAC_HORSE	Q28389 equus caball
44	260	12.1	345	1	CARD_PIG	P00795 sus scrofa
45	255	11.8	496	1	ASPR_ORYSA	P42211 oryza sativ

## ALIGNMENTS

RESULT 1			
BACE_HUMAN	STANDARD:	PRT:	501 AA.
ID	P56817: Q9UT5;		
AC	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)		
DE	(Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl		
DE	protease 2) (Asp 2) (Asp2) (Membrane-associated aspartic protease 2)		
DE	(Memapsin-2).		
GN	BACE OR BACE1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM A).		
RC	TISSUE=Brain;		
RX	MEDLINE=20002972; PubMed=10531052;		
RA	Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,		
RA	Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,		
RA	Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,		
RA	Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,		
RA	Treanor J., Rogers G., Citron M.;		
RT	"Beta-secretase cleavage of Alzheimer's amyloid precursor protein by		
RT	the transmembrane aspartic protease BACE.";		
RL	Science 286:735-741(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND		
RC	CHARACTERIZATION.		
RX	TISSUE=Brain;		
RA	MEDLINE=20057171; PubMed=10591214;		
RA	Sinha S., Anderson J.P., Barbour R., Basl G.S., Caccavello R.,		
RA	David D., Doan M., Dovey H.F., Fridgen N., Hong J., Jacobson-Croak K.,		
RA	Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,		
RA	Tatsuno G., Tung J., Schenk D., Seubert P., Suomensaar S.M., Wang S.,		
RA	Walker D., Zhao J., McConlogue L., Varghese J.;		
RT	"Purification and cloning of amyloid precursor protein beta-secretase		
RT	from human brain.";		
RL	Nature 402:537-540(1999).		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM A).		
RX	MEDLINE=20057170; PubMed=10591213;		
RA	Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,		
RA	Brashner J.R., Strutham N.C., Mathews W.R., Buhl A.E., Carter D.B.,		
RA	Thammaselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.;		
RT	"Membrane-anchored aspartyl protease with Alzheimer's disease beta-		
RT	secretase activity ";		
RL	Nature 402:533-537(1999).		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM A).		
RX	MEDLINE=20030166; PubMed=10561122;		
RA	Hussein I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,		
RA	Chapman C., Glover I.S., Murphy K.E., Southan C.D., Ryan D.M.,		
RA	Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;		

RT "identification of a novel aspartic proteinase (Asp 2) as beta-secrese.";  
 RL Mol. Cell. Neurosci. 14:419-427(1999).  
 RN [5].  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Brain, and Pancreas;  
 RA Michel B., De Pietri Tonelli D., Zaccchetti D., Keller P.;  
 RT "New beta-site APP cleaving enzyme isoform (BACE-1B) obtained from human brain and pancreas";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6].  
 RP SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.  
 RX MEDLINE=2014060; PubMed=10677483;  
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;  
 RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP, LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A/BACE-1A (shown here) and B/BACE-1B; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EURKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF190725; AAF04142.1; -  
 DR EMBL: AF201468; AAF18982.1; -  
 DR EMBL: AF200343; AAF17079.1; -  
 DR EMBL: AF204943; AAF26367.1; -  
 DR EMBL: AF338816; AAK38374.1; -  
 DR EMBL: AF200193; AAF13715.1; -  
 DR HSSP: P32329; 1YPS.  
 DR MIM: 604252; -  
 DR InterPro: IPR001969; Asp.-protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp; 3.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP-PROTEASE; 1.  
 KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;  
 FT Signal; Alternative splicing.  
 FT SIGNAL 1 21  
 FT PROPEP 22 45  
 FT CHAIN 46 501  
 FT DOMAIN 22 457  
 FT TRANSLEM 458 478  
 FT DOMAIN 479 501  
 FT ACT\_SITE 93 93  
 FT ACT\_SITE 289 289  
 FT CARBOHYD 153 153  
 FT CARBOHYD 172 172  
 FT CARBOHYD 223 223  
 FT CARBOHYD 354 354  
 FT VARSPLIC 190 214  
 FT VARSPLIC 377CE4C824ACEF05 CRC64;  
 SQ SEQUENCE 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;  
 Query Match 100.0%; Score 2156; DB 1; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 1,3e-171;  
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETDEEPEERGRSGFVEMVDNLRGKSGGYVEMTVGSPQTLNIIYDTGSSNFANGAP 60  
 DB 46 ETDEEPEERGRSGFVEMVDNLRGKSGGYVEMTVGSPQTLNIIYDTGSSNFANGAP 105  
 QY 61 HPLHLRYQQLSTYRDLAKGYVPTGKMGELGTDIVSIPHGPNYVRANIAITE 120  
 DB 106 HPLHLRYQQLSTYRDLAKGYVPTGKMGELGTDIVSIPHGPNYVRANIAITE 165  
 QY 121 SDRFFINGSWMEGLGLAVAEIARPPDLSLPPFDSLVKQTHVNLFSLOLCGAFPLNOS 180  
 DB 166 SDRFFINGSWMEGLGLAVAEIARPPDLSLPPFDSLVKQTHVNLFSLOLCGAFPLNOS 225  
 QY 181 EVIASVGSMTIGGIDHSLYTGSIMTYPIREMYEYIYRVEINSGDLMCKEKNYDK 240  
 DB 226 EVIASVGSMTIGGIDHSLYTGSIMTYPIREMYEYIYRVEINSGDLMCKEKNYDK 285  
 QY 241 SIYVSGTTLRLPKFVEAAVKSIAASSTKEKPDGFWLGEOLVCMQAGTTPMNIPEVIS 300  
 DB 286 SIYVSGTTLRLPKFVEAAVKSIAASSTKEKPDGFWLGEOLVCMQAGTTPMNIPEVIS 345  
 QY 301 LYNGEVTNOSFRITLLPQOYLREVEDVATISODDCYFAISQSTGYMGAIVMEGFYV 360  
 DB 346 LYNGEVTNOSFRITLLPQOYLREVEDVATISODDCYFAISQSTGYMGAIVMEGFYV 405  
 QY 361 FDRARKRIGFAVSACHVDEPRTAAVSGPVTLMEDCGYNIPTQDE 407  
 DB 406 FDRARKRIGFAVSACHVDEPRTAAVSGPVTLMEDCGYNIPTQDE 452  
 RESULT 2  
 BACE\_RAT 501 AA.  
 ID BACE\_RAT STANDARD: PRT: 501 AA.  
 AC P56819;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
 DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
 DE protease 2) (Asp 2) (Asp2) (Membrane-associated aspartic protease 2)  
 DE (Memapsin-2).  
 GN GN  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20002972; PubMed=10531052;  
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
 RA Fisher S., Fuller J., Curren E., Burgess T., Louis J.-C., Collins F.,  
 RA Treanor J., Rogers G., Citron M.;  
 RA "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE";  
 RT Science 286:735-741(1999).  
 RL  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP, LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASED BY GAMMA-SECRETASE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EURKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).

CC EMBL: AF190727; AAF04143.1; -  
 DR HSSP: P32329; 1YPS.  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp. 3.  
 DR PRINTS: PR00792; Pepsin.  
 DR PROSITE: PS00141; Asp\_PROTEASE; 1.  
 DR HydroLase: Aspartyl protease; Glycoprotein; zymogen; Transmembrane;  
 signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT PROPEP 22 45 POTENTIAL.  
 FT CHAIN 46 501 BETA-SECRETASE.  
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 458 478 POTENTIAL.  
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).  
 FT ACT\_SITE 93 93 BY SIMILARITY.  
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 501 AA; 55806 MW; 2484458C8B87DE3 CRC64;

Query Match Best Local Similarity 98.9%; Score 2132; DB 1; Length 501;  
 Matches 400; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ETDEPEPEGRGRSFEVEMVNLGRKSGQGYVEMTGTSPQTNTILVDTGSSNFANGAP 60  
 DB 46 ETDEPEPEGRGRSFEVEMVNLGRKSGQGYVEMTGTSPQTNTILVDTGSSNFANGAP 105  
 QY 61 HPLRLRYQQLSTTRDLKRGVYVYTGKMEGLGTLVSLPHGPNTVANAIAIE 120  
 DB 106 HPLRLRYQQLSTTRDLKRGVYVYTGKMEGLGTLVSLPHGPNTVANAIAIE 165  
 QY 121 SDRFFNGSMWEGILGLAVAEIARPDLSLEPPDLSVKQTHPNFSLDLCAGPEPLNOS 180  
 DB 166 SDRFFNGSMWEGILGLAVAEIARPDLSLEPPDLSVKQTHPNFSLDLCAGPEPLNOS 225  
 QY 181 EVLASVSGMTIGGIDSLYTGSLWTPPIRBMVYEVIIIVREINGQDLMCKCKEYNDK 240  
 DB 226 EVLASVSGMTIGGIDSLYTGSLWTPPIRBMVYEVIIIVREINGQDLMCKCKEYNDK 285  
 QY 241 SIVDSSTNLRLPKKVEAAVKSIRAASTKEKPPGPFMLGEOVLCWQAGTTPWNIPIVIS 300  
 DB 286 SIVDSSTNLRLPKKVEAAVKSIRAASTKEKPPGPFMLGEOVLCWQAGTTPWNIPIVIS 345  
 QY 301 LYLGEVTVNQSFRTITLPQOYLRPVEDVATSQDDCYKFAISOSSTGTWAGVIMGEFYV 360  
 DB 346 LYLGEVTVNQSFRTITLPQOYLRPVEDVATSQDDCYKFAISOSSTGTWAGVIMGEFYV 405  
 QY 361 FPRARRKRGFAVSAHVDEFTAVVEGPFYTLDMEDCGYNIPQIDE 407  
 DB 406 FPRARRKRGFAVSAHVDEFTAVVEGPFYTLDMEDCGYNIPQIDE 452

RESULT 3  
 BACE\_MOUSE STANDARD: PRT; 501 AA.  
 AC P56818;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)  
 DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl  
 DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)  
 DE (Memapsin-2).  
 GN BACE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20002972; PubMed=10531052;  
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,  
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,  
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,  
 RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,  
 RA Treanor J., Rogers G., Citron M.,  
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by  
 the transmembrane aspartic protease BACE.";  
 RL Science 286:735-741(1999).  
 RN [2]  
 RP REVISIONS TO 6 AND 81-87.  
 RA Bennett B.D., Vassar R., Citron M.,  
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057170; PubMed=10591213;  
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Torry M.C., Pauley A.M.,  
 RA Braslier J.R., Strahman N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
 RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;  
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease  
 beta-secretase activity.";  
 RL Nature 402:533-537(1999).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE  
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF  
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,  
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED  
 CC FRAGMENT APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: BRAIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EKURAROTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 DR EMBL: AF190726; AAF04143.2; -  
 DR EMBL: AF200346; AAF17082.1; -  
 DR HSSP: P36272; 1AM5.  
 DR MGD: MGI:1346542; Bace.  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp. 1.  
 DR PRINTS: PR00792; Pepsin.  
 DR PROSITE: PS00141; Asp\_PROTEASE; 1.  
 DR HydroLase: Aspartyl protease; Glycoprotein; zymogen; Transmembrane;  
 DR signal.  
 KW SIGNAL 1 21 POTENTIAL.  
 KW PROPEP 22 45 POTENTIAL.  
 KW CHAIN 46 501 BETA-SECRETASE.  
 KW DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).  
 KW TRANSMEM 458 478 POTENTIAL.  
 KW DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).  
 KW ACT\_SITE 93 93 BY SIMILARITY.  
 KW CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 KW CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 501 AA; 55747 MW; C085A0131458474E CRC64;

Query Match 98.9%; Score 2132; DB 1; Length 501;

Best Local Similarity 98.3%; Pred. No. 1.3e-169;  
Matches 400; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 ETDEPEPEGRSGSEFMVDNLKSGGQGYVEMTVSSPQTLNIIIVDSSNFAVGAAP 60  
 DB 46 ETDESEPEGRSGSEFMVDNLKSGGQGYVEMTVSSPQTLNIIIVDSSNFAVGAAP 105  
 OY 61 HPELHRYQRLSTYRDLRKGYYVPTQCKMEGELGTDLVSIIPHGNVTVRANIAITE 120  
 DB 106 HPELHRYQRLSTYRDLRKGYYVPTQCKMEGELGTDLVSIIPHGNVTVRANIAITE 165  
 OY 121 SDKPFINSNMEGILGLAVAEIARPDLSLEPFDSLVKQTHVPLVFSIQCGAFPLNOS 180  
 DB 166 SDKPFINSNMEGILGLAVAEIARPDLSLEPFDSLVKQTHVPLVFSIQCGAFPLNOS 225  
 OY 181 EYLVASVSGSMIIGIDHSLYGLSLMYPIRREMYEVIIVAEINOGDKMDCKEYNDK 240  
 DB 226 EYLVASVSGSMIIGIDHSLYGLSLMYPIRREMYEVIIVAEINOGDKMDCKEYNDK 285  
 OY 241 SYVDSGTNLRLPKVFPAAYKSIKAASSTKPFDPGFWLGEOLVCMQAGTTPWNIPIYIS 300  
 DB 286 SYVDSGTNLRLPKVFPAAYKSIKAASSTKPFDPGFWLGEOLVCMQAGTTPWNIPIYIS 345  
 OY 301 LYLMGEVNTQSFRTITLPOQYLRPEVDATSDDCYKFAISGSGTYMGAVIMEGYV 360  
 DB 346 LYLMGEVNTQSFRTITLPOQYLRPEVDATSDDCYKFAISGSGTYMGAVIMEGYV 405  
 OY 361 PPRARRIGFAVSACHVHEPRTAAVEGPFVLDMEDCGYNIPQDE 407  
 DB 406 PPRARRIGFAVSACHVHEPRTAAVEGPFVLDMEDCGYNIPQDE 452

RESULT 4  
 BAE2 HUMAN STANDARD; PRT; 518 AA.

AC 09Y520; OGUUT6;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving enzyme 2) (Aspartyl protease 1) (Asp 1) (Asp1) (Membrane-associated aspartic protease 1) (Memapsin-1).  
 GN BACE2 OR ASP21.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBI\_TaxID=9606;  
 OX 11  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=20051170; PubMed=10591213;  
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,  
 RA Brashler J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,  
 RA Tomasselli A.G., Parodi L.A., Heinrichson R.L., Gurney M.E.,  
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease  
 RT beta-secretase activity.";  
 RT Nature 402:533-537(1999).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Bone marrow;  
 RC Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,  
 RA Giese K.;  
 RT "Identification of a novel aspartic-like protease differentially  
 RT expressed in human breast cancer cell lines";  
 RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Acciarini M.P., Fumagalli P., Ottolenghi S., Taramelli R.;  
 RT "Cloning of a gene from chromosome 21 Down region encoding a potential  
 RT transmembrane aspartyl protease.";  
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Solans A., Estivill X., de la Luna S.;

RT "Cloning of a novel mammalian aspartyl protease.";  
 RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20030166; PubMed=10561122;  
 RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,  
 RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,  
 RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;  
 RT "Identification of a novel aspartic proteinase (asp 2) as  
 RT beta-secretase.";  
 RT Mol. Cell. Neurosci. 14:419-427(1999).  
 RL [6]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=2014406; PubMed=10677483;  
 RX Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;  
 RA "Human aspartic protease memapsin 2 cleaves the beta-secretase site of  
 RT beta-amyloid precursor protein.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20289799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Soeda E.,  
 RA Park H.-S., Toyoda A., Ichii K., Totoki Y., Choi D.-K., Polley A.,  
 RA Onki M., Takagi T., Sakai Y., Taudien S., Blechschmidt K., Polley A.,  
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
 RA Minoshima S., Shimizu N., Nordstieck G., Hornischer K., Brandt P.,  
 RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,  
 RA Kanster J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,  
 RA Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 RA Lehnach H., Reinhardt R., Yaspo M.-L.;  
 RT "The DNA sequence of human chromosome 21.";  
 RL Nature 405:311-319(2000).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF200342; AAF17078.1;  
 DR EMBL: AF117892; AAD45240.1;  
 DR EMBL: AF050171; AAD45963.1;  
 DR EMBL: AF178532; AAF29494.1;  
 DR EMBL: AF204944; AAF26368.1;  
 DR EMBL: AF200192; AAF13714.1;  
 DR EMBL: AL163284; CAB90458.1;  
 DR EMBL: AL163285; CAB90554.1;  
 DR HSSP: P00797; ZREN.  
 DR MIM: 605668;  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp. 3.  
 DR PRINTS: PS00141; ASP\_PROTEASE. 2.  
 DR PROSITE: PS00141; ASP\_PROTEASE. 2.  
 KW Hydroxylase; Aspartyl protease; Glycoprotein; zymogen; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 20  
 FT PROPEP 21 518  
 FT CHAIN ? 518  
 FT DOMAIN 21 473  
 FT DOMAIN 474 494  
 FT DOMAIN 495 518  
 FT DOMAIN 110 110  
 FT ACT\_SITE 303 303  
 FT ACT\_SITE 170 170  
 FT CARBOHYD 170 170  
 FT POTENTIAL.  
 FT POTENTIAL.  
 FT BETA\_SECRETASE 2.  
 FT EXTRACELLULAR (POTENTIAL).  
 FT CYTOPLASMIC (POTENTIAL).  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).



Query Match	Best Local Similarity	Matches	Score	DB 1:	Length	DB 2:	Score	DB 3:	Length	DB 4:	Score	DB 5:	Length	DB 6:	Score	DB 7:	Length	DB 8:	Score	DB 9:	Length	DB 10:	Score	DB 11:	Length	DB 12:	Score	DB 13:	Length	DB 14:	Score	DB 15:	Length	DB 16:	Score	DB 17:	Length	DB 18:	Score	DB 19:	Length	DB 20:	Score	DB 21:	Length	DB 22:	Score	DB 23:	Length	DB 24:	Score	DB 25:	Length	DB 26:	Score	DB 27:	Length	DB 28:	Score	DB 29:	Length	DB 30:	Score	DB 31:	Length	DB 32:	Score	DB 33:	Length	DB 34:	Score	DB 35:	Length	DB 36:	Score	DB 37:	Length	DB 38:	Score	DB 39:	Length	DB 40:	Score	DB 41:	Length	DB 42:	Score	DB 43:	Length	DB 44:	Score	DB 45:	Length	DB 46:	Score	DB 47:	Length	DB 48:	Score	DB 49:	Length	DB 50:	Score	DB 51:	Length	DB 52:	Score	DB 53:	Length	DB 54:	Score	DB 55:	Length	DB 56:	Score	DB 57:	Length	DB 58:	Score	DB 59:	Length	DB 60:	Score	DB 61:	Length	DB 62:	Score	DB 63:	Length	DB 64:	Score	DB 65:	Length	DB 66:	Score	DB 67:	Length	DB 68:	Score	DB 69:	Length	DB 70:	Score	DB 71:	Length	DB 72:	Score	DB 73:	Length	DB 74:	Score	DB 75:	Length	DB 76:	Score	DB 77:	Length	DB 78:	Score	DB 79:	Length	DB 80:	Score	DB 81:	Length	DB 82:	Score	DB 83:	Length	DB 84:	Score	DB 85:	Length	DB 86:	Score	DB 87:	Length	DB 88:	Score	DB 89:	Length	DB 90:	Score	DB 91:	Length	DB 92:	Score	DB 93:	Length	DB 94:	Score	DB 95:	Length	DB 96:	Score	DB 97:	Length	DB 98:	Score	DB 99:	Length	DB 100:	Score	DB 101:	Length	DB 102:	Score	DB 103:	Length	DB 104:	Score	DB 105:	Length	DB 106:	Score	DB 107:	Length	DB 108:	Score	DB 109:	Length	DB 110:	Score	DB 111:	Length	DB 112:	Score	DB 113:	Length	DB 114:	Score	DB 115:	Length	DB 116:	Score	DB 117:	Length	DB 118:	Score	DB 119:	Length	DB 120:	Score	DB 121:	Length	DB 122:	Score	DB 123:	Length	DB 124:	Score	DB 125:	Length	DB 126:	Score	DB 127:	Length	DB 128:	Score	DB 129:	Length	DB 130:	Score	DB 131:	Length	DB 132:	Score	DB 133:	Length	DB 134:	Score	DB 135:	Length	DB 136:	Score	DB 137:	Length	DB 138:	Score	DB 139:	Length	DB 140:	Score	DB 141:	Length	DB 142:	Score	DB 143:	Length	DB 144:	Score	DB 145:	Length	DB 146:	Score	DB 147:	Length	DB 148:	Score	DB 149:	Length	DB 150:	Score	DB 151:	Length	DB 152:	Score	DB 153:	Length	DB 154:	Score	DB 155:	Length	DB 156:	Score	DB 157:	Length	DB 158:	Score	DB 159:	Length	DB 160:	Score	DB 161:	Length	DB 162:	Score	DB 163:	Length	DB 164:	Score	DB 165:	Length	DB 166:	Score	DB 167:	Length	DB 168:	Score	DB 169:	Length	DB 170:	Score	DB 171:	Length	DB 172:	Score	DB 173:	Length	DB 174:	Score	DB 175:	Length	DB 176:	Score	DB 177:	Length	DB 178:	Score	DB 179:	Length	DB 180:	Score	DB 181:	Length	DB 182:	Score	DB 183:	Length	DB 184:	Score	DB 185:	Length	DB 186:	Score	DB 187:	Length	DB 188:	Score	DB 189:	Length	DB 190:	Score	DB 191:	Length	DB 192:	Score	DB 193:	Length	DB 194:	Score	DB 195:	Length	DB 196:	Score	DB 197:	Length	DB 198:	Score	DB 199:	Length	DB 200:	Score	DB 201:	Length	DB 202:	Score	DB 203:	Length	DB 204:	Score	DB 205:	Length	DB 206:	Score	DB 207:	Length	DB 208:	Score	DB 209:	Length	DB 210:	Score	DB 211:	Length	DB 212:	Score	DB 213:	Length	DB 214:	Score	DB 215:	Length	DB 216:	Score	DB 217:	Length	DB 218:	Score	DB 219:	Length	DB 220:	Score	DB 221:	Length	DB 222:	Score	DB 223:	Length	DB 224:	Score	DB 225:	Length	DB 226:	Score	DB 227:	Length	DB 228:	Score	DB 229:	Length	DB 230:	Score	DB 231:	Length	
-------------	-----------------------	---------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	--



AC P27678:  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Pepsin A-4 precursor (EC 3.4.23.1) (Pepsin I/II).  
 GN PCA.  
 OS Macaca fuscata fuscata (Japanese macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 NCBI\_TaxID=9543;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.  
 RC TISSUE=gastric mucosa.  
 RX MEDLINE=92037645; PubMed=1935977;  
 RA Kageyama T., Tanabe K., Kolwai O.;  
 RT "Development-dependent expression of isozymes of monkey  
 RT pepsinogens and structural differences between them."  
 RL Eur. J. Biochem. 202:205-215(1991).  
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
 CC HORMONES AND RELATED SUBSTANCES.  
 CC -1- MISCELLANEOUS: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING  
 CC PEPSIN AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA  
 CC ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC  
 CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to: license@isb-sib.ch).  
 CC -----  
 DR EMBL: X59753; CAA2425.1; -.  
 DR PIR: S16065; S16065.  
 DR PIR: S19682; S19682.  
 DR HSSP: P00790; IPSN.  
 DR MEROPS: A01.001; -.  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001961; Laminin\_G.  
 DR Pfam: PF00026; Pepsin.  
 DR PRINTS: PS00141; ASP\_PROTEASE; 2.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydroxylase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
 KW Zymogen; Multigene family; Signal; Glycoprotein.  
 FT SIGNAL 1 15  
 FT PROPEP 16 38 ACTIVATION PEPTIDE.  
 FT PROPEP 39 62 ACTIVATION PEPTIDE.  
 FT CHAIN 63 388 PEPSIN A-4.  
 FT ACT\_SITE 94 94 BY SIMILARITY.  
 FT ACT\_SITE 277 277 BY SIMILARITY.  
 FT DISULFID 107 112 BY SIMILARITY.  
 FT DISULFID 268 272 BY SIMILARITY.  
 FT DISULFID 311 344 BY SIMILARITY.  
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .).  
 SQ SEQUENCE 388 AA; 4195 MW; A2923AB1F7FCD89 CRC64;

Query Match 14.0%; Score 302.5; DB 1; Length 388;  
 Best Local Similarity 27.6%; Pred. No. 1e-17;  
 Matches 107; Conservative 64; Mismatches 133; Indels 83; Gaps 17;

OY 3 DEEPEERGRSFEVMDNLRKSGGCGYVMTVGSPPQTLNLTIVTGGSSFAVGAAPHP 62  
 DB 64 DEQPLE-----NYDVF-----EYFGTIGTGPANQNTVEVDGSSNTLWV---PSV 105

OY 63 FL-----HRYQRLSTYRDLKRVVPTQGWSEGLTDLVSIPIHGNTVRANI 115  
 DB 106 YCYSIACMDHNFENPDSSSTRATSKTWSITFGTSMGILGDIYHVK---CGISDNTQI 162  
 OY 116 AAITESDK-FFINGSMWEIILGLAVAEIARPPDSELEPPFDSLYKQTH-PNIFSQLCGA 173  
 DB 163 FGLSETEPGEFLFXAPFDGILALAPSIS--SSGATPVPDNTMNRLVSQDLFSYLSAD 220  
 OY 174 GEPLOSSEVLASVSGSMIIGIDHSLYTGSLWYTPIREWYEVYIIVVEINGODL--KM 231  
 DB 221 ---DQS-----GSVIFEGIDSSYTGSLWNVVPSVGVQWQISVDITNNGKTIACAK 270  
 OY 232 DCKEYNKDSIYDSTTNLRPKKVFEAIVSIRKASSTKFPDGFUIGEDLV-CWQAGT 290  
 DB 271 GC-----QATVDITSLTGTSPSIANIQSDIGASSENS-----GEMVSSCAISS 316  
 OY 291 TPWNIFPYISLYIMEVYNOSFRITLPQY-LRPVEVATNSODDCYK-----FAISQSS 344  
 DB 317 LEDIVE-----TINGVQYPLPRAIILIOSGSCISFGQMDVPIESG 358  
 OY 345 TGTVGAVIMEGFYVDFRARRIGEFA 371  
 DB 359 ELMILGDVEIRQYFTVDFRANNOVGLA 385

RESULT 9  
 PEPA\_CHICK STANDARD; PRT; 367 AA.  
 AC P00793;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Pepsin A precursor (EC 3.4.23.1).  
 OS Gallus gallus (chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE=84004412; PubMed=6617663;  
 RA Baudys M., Kostka V.;  
 RT "Covalent structure of chicken pepsinogen."  
 RL Eur. J. Biochem. 136:89-99(1983).  
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC PIR: A00984; PECH.  
 DR HSSP: P00794; ACMS.  
 DR MEROPS: A01.001; -.  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; Asp; 1.  
 DR PRINTS: PR00792; PEPsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Hydroxylase; Aspartyl protease; Digestion; Stomach; Zymogen;  
 KW Glycoprotein; Gastric juice.  
 FT PROPEP 1 42 ACTIVATION PEPTIDE.  
 FT CHAIN 43 367 PEPSIN A.  
 FT ACT\_SITE 77 77  
 FT ACT\_SITE 260 260  
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).  
 FT DISULFID 90 95  
 FT DISULFID 251 255  
 FT DISULFID 290 323  
 SQ SEQUENCE 367 AA; 40431 MW; 0C547E7ED8F5B341 CRC64;

Query Match 14.0%; Score 302; DB 1; Length 367;

Best Local Similarity 24.0%; Pred. No. 1e-17;  
Matches 88; Conservative 69; Mismatches 125; Indels 84; Gaps 13;

```

OY 30 YVEMTVGSPQPTINILVDGSSNFVANGAPHPRL-----HRYGROLSSTYRDLKRG 82
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 YGTVISIGTPOODDSVIFDTGSSNLMWV---PSYCKSSACSNHRRFPSSSTIYVSTNET 115
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 83 VVYVYTGKMGEGELGTDLSIPHGPNVYVRANIAITESDK-FPINGSNMEGILGLAYAE 141
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 116 VYVAVGTSMSGIIGIDYVAV---SSIDVQNOIGLSTEGSEFFYCNFGDIGLAFPS 172
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 142 IARPDLSLEPFDSLKYQTHV-PNLFSLQLCGAGFPPLNQSVLASVGSMTIGGIDISLY 200
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 173 IS--SSGATPVEDNMMSOHLVAODLFFSYLSKDG-----ETGSEVLEGGIDPYTT 220
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 201 TGSIMYTPIREMYEYVIVEIYVEINGDLK--MDCKEYNDKSIYDGGTNLRPKVFE 258
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 221 TKGIWVPLSSETWQITMDRYVGNKRYVACEFTC-----QALVDIGTSLVMDQAGYN 274
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 259 AAVSIRKASASTE-----KFPDGFMLGEOLVCMQACTPWNIFFPVISLYLMEVNTQS 311
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 275 RIKNGLVSSDDEISCDISKLPD-----VTFHNGHA--- 307
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 312 FRITTLPOOYLPRVEDVATSDQDCKKFAISQSGT-----VMGAVIMEGFYVDFRAR 365
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 308 -----FTLPASAVYVLNEDGSCMLGFENMGTFTEIGEOMITLDVFIREYVIFEDKAN 358
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 366 KRIGFA 371
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 NKVGLS 364
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 10
CATE HUMAN STANDARD; PRI: 396 AA.
ID P14091:
AC 01-JAN-1990 (rel. 13, Created)
DT 01-JAN-1990 (rel. 13, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Cathepsin E precursor (EC 3.4.23.34).
GN CTSE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89380302; PubMed=2674141;
RA Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;
RT "Human gastric cathepsin E. Predicted sequence, localization to
RT chromosome 1, and sequence homology with other aspartic
RT proteinases."
RT J. Biol. Chem. 264:16748-16753(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112877; PubMed=1370478;
RA Azuma T., Liu W.G., Vander laan D.J., Bowcock A.M., Taggart R.T.;
RT "Human gastric cathepsin E gene. Multiple transcripts result from
RT alternative polyadenylation of the primary transcripts of a single
RT gene locus at 1q31-q32."
RL J. Biol. Chem. 267:1609-1614(1992).
RN [3]
RP SEQUENCE FROM N.A.
RA Tanneil P.J., Kay J.;
RT "Human procathepsin E."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.
RX MEDLINE=90241267; PubMed=2334440;
RA Athanada S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;
RT "Structural evidence for two isozytic forms and the carbohydrate
RT attachment site of human gastric cathepsin E."
RL Biochem. Biophys. Res. Commun. 168:878-885(1990).

```

```

CC - FUNCTION: DUE OF ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC - CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
CC specificity.
CC - SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
Cc EMBL; M84424; AAA52300.1; -
Dr EMBL; M84413; AAA52300.1; JOINED.
Dr EMBL; M84417; AAA52300.1; JOINED.
Dr EMBL; M84418; AAA52300.1; JOINED.
Dr EMBL; M84419; AAA52300.1; JOINED.
Dr EMBL; M84420; AAA52300.1; JOINED.
Dr EMBL; M84421; AAA52300.1; JOINED.
Dr EMBL; M84422; AAA52300.1; JOINED.
Dr EMBL; J05036; AAA52130.1; -
Dr EMBL; A1250717; CAB82850.1; -
Dr PIR; A34401; A34401.
Dr PIR; A34643; A34643.
Dr PIR; A42038; A42038.
Dr HSSP; P00794; 4CMS.
Dr MEROPS; A01.010; -.
Dr MIM; 116890; -.
Dr InterPro; IPR001969; Asp_protease.
Dr InterPro; IPR001461; Pepsin.
Dr Pfam; PF00026; pepsin.
Dr PRINTS; PR00792; PEPsin.
Dr PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; signal.
FT SIGNAL 1 17
FT PROPEP 18 53
FT CHAIN 54 396
FT MOD_RES 18 18
FT ACT_SITE 96 96
FT ACT_SITE 281 281
FT DISULFID 60 60
FT DISULFID 109 114
FT DISULFID 272 276
FT DISULFID 314 351
FT CARBOHYD 90 90
FT CARBOHYD 220 220
FT CARBOHYD 333 333
SQ SEQUENCE 396 AA; 42793 MW; 40B643C5FB01521E CAC64;

Query Match 13.9%; Score 299.5; DB 1; Length 396;
Best Local Similarity 25.9%; Pred. No. 1e-17;
Matches 100; Conservative 65; Mismatches 148; Indels 73; Gaps 16;

OY 3 DEEPPEPGRGSEFVENDLKRKSGGGYVEMTVGSPQPTINILVDGSSNFAVA----- 58
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 63 DQSAKEP-----LINYLD-----MEYFGTISISPPONTFVIRHDTGSSNLMWVSVCYT 110
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 59 APHPFLHRYQRLSTYVDLKRKYVVPYPTQCKMGEELGTDLSIPHGPNVYVRANIAI 118
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 SPACKTHSRFOPSSSTYQPOGSEFISIQGTSLGSIITADYVAV-EGLTIVGQGFGESEV 169
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 119 TESDKFPINGSNMEGILGLAYAEIARPDLSLEPFDSLKYQTHVPLNLFSLQLCGAGFPPLN 178
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 TEGGQGFVD-AEFDGILIGLGYPSLA--VGGYVPVGFNMMAQ-----NLVLDLPMFSYVMSN 222
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 179 QSEVLASVGSMTIGGIDLSLYTGSIMYTPIREMYEYVIVEIYVEINGDLKMDCKEYVAV 238
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 223 PE--GGAGSELLIFGVDHSHFSGSLNWVAVTVTKQAVQIALDNIQVG--TVWFSE--G 275
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

QY 239 DKSTVDSSTNLRRLPKRYEAAVSKAASSTKFFPDGFWLGVQWQAGTTPMNIFFV 298  
 Db 276 COATVDTGTSITGSPDKIKLOLNAIGAAFP-----VGEYAVE-----CANLNMWD 322  
 QY 299 ISLYLMEGVNTQNSRITILPQOYLPRPEADVATSDDDCKFAISQSSFG----- 346  
 Db 323 VTFITNG-----VPTLSPFAY--TLLDFVDGMQFC-----SSGFGGLDIHPAPG 366  
 QY 347 -TVMGAVIMEGFVVPFRRARRIGFA 371  
 Db 367 LMLGDVFIRQFYSVFPDRGNRVGLA 392

RESULT 11  
 PEPE\_CHICK STANDARD; PRT; 383 AA.  
 AC P16476;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Embryonic pepsinogen precursor (EC 3.4.23.-).  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88227903; PubMed=3131317;  
 RX Hayashi K., Agata K., Mochi M., Yasugi S., Eguchi G., Mizuno T.;  
 RT "Molecular cloning and the nucleotide sequence of cDNA for embryonic  
 RT chicken pepsinogen: phylogenetic relationship with prochymosin";  
 RL J. Biochem. 103:290-296(1986).  
 CC -1- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC  
 CC PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D00215; BAA00153.1; -.  
 DR PIR: A41443; A41443.  
 DR HSSP: P00794; 4CMS.  
 DR MEROPS: A01.028; -.  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp.1.  
 DR PRINTS: PRO0792; Pepsin.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 DR KMW Hydroxylase; Aspartyl protease; Digestion; Signal; Glycoprotein.  
 FT SIGNAL 1 16  
 FT CHAIN 17 383  
 FT ACT\_SITE 94 94  
 FT ACT\_SITE 276 276  
 FT DISULFID 107 112  
 FT DISULFID 267 271  
 FT DISULFID 310 344  
 FT CARBOHYD 132 132  
 FT CARBOHYD 204 204  
 FT CARBOHYD 309 309  
 FT CARBOHYD 350 350  
 FT VARIANT 51 51  
 FT SEQUENCE 383 AA; 41719 MW; 1642796871611F54 CRC64;

Query Match 13.8%; Score 298.5; DB 1; Length 383;

Best Local Similarity 25.2%; Pred. No. 2.2e-17;  
 Matches 90; Conservative 75; Mismatches 125; Indels 67; Gaps 14;

QY 30 YVEMTVGSPQPTLNLVPTGSSNFAVGA-----APRFLHRYKXQRLSTYDLKQYV 85  
 Db 76 YVGTISIGTRPDQFVYVPTDGSNLMVPSVCTSPACOSHOMENPSQSTYSTGQNSI 135  
 QY 86 PYTGKWESEGLGDLVSIHPGPNVYRANIAMITESDKFFINGSMWEGILGLVAEIRP 145  
 Db 136 HGTGDMEGTGVGDVTVASLMDTNQLFELST-SEPGQFFV-VAFDGLIGLGYSLA- 192  
 QY 146 DDLPEPDSLVKQTHV-PNLFSLOLCGAGFPLNOSVLAISGSMITIGDHSLYTSSL 204  
 Db 193 -DGITVFEDNMWNESSLLEONLFVYLS-----REPMSVYVGGIDSEYFGSI 240  
 QY 205 WYTPIRREWYEVYIVRVINGQDL--KMDCKEYVDKSYVSGTNTLRPKRYEAAV 262  
 Db 241 NMIPVSYQGYWQISMDSTIVNKOELACSSG-----QALIDTGLVAGPSDINDIOS 294  
 QY 263 SKAASSTKFPDGFVLGVQWQAGTTPMNIFFVSL--YLMGEVTVQSFRTILP 318  
 Db 295 AVGANQNT-----YGEYSV-----NCHITLMPDVVFYIGI----- 326  
 QY 319 QOYLPRVEDVA---TSDDCKFAISQSSSTGVVGVAVIMEGFVVPFRRARRIGFA 371  
 Db 327 -QY--PVRLATTEQNGQGTGCMSSFQNSADLMLGDVFIRYYSIFDRANNRVGLA 380

RESULT 12  
 PEP2\_RABIT STANDARD; PRT; 387 AA.  
 ID PEP2\_RABIT  
 AC P27821;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Pepsin II-2/3 precursor (EC 3.4.23.1) (Pepsin A).  
 OS Oryctolagus cuniculus (Rabbit).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 CC NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91009127; PubMed=2129536;  
 RX Kageyama T., Tanabe K., Koizumi O.;  
 RT "Structure and development of rabbit pepsinogens. Stage-specific  
 RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and  
 RT gene expression during development.";  
 RL J. Biol. Chem. 265:17031-17038(1990).  
 CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS  
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
 CC ALSO CLEAVED TO SOME EXTENT.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
 CC -1- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE  
 CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.  
 CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY  
 CC HORMONES AND RELATED SUBSTANCES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M59225; AAA85369.1; -.  
 DR PIR: C38302; C38302.  
 DR HSSP: P00790; IPSN.  
 DR MEROPS: A01.001; -.  
 DR InterPro: IPR001969; Asp\_protease.  
 DR InterPro: IPR001461; Pepsin.



```

DB      334 Y--TLDEVDGMQVC-----STGEESLEIQPPAGPLMLGDVFIROFYAVFEDRGNMR 383
QY      368 IGFA 371
DB      384 VGLA 387

RESULT 14
CATD_HUMAN
ID      CATD_HUMAN      STANDARD:      PRT:      412 AA.
AC      P07339;
DT      01-APR-1988 (rel. 07, Created)
DT      01-APR-1988 (rel. 07, Last sequence update)
DT      01-MAR-2002 (rel. 41, Last annotation update)
DE      Cathepsin D precursor (EC 3.4.23.5).
GN      CTSD.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=85270436; PubMed=3927292;
RA      Faust P.L., Kornfeld S., Chirgwin J.M.;
RT      "Cloning and sequence analysis of cDNA for human cathepsin D.";
RL      Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).
RN      (2)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87231068; PubMed=3588310;
RA      Westley B.R., May F.E.B.;
RT      "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive
RL      human breast cancer cells.";
RN      (3)
RP      Nucleic Acids Res. 15:3773-3786(1987).
RN      (4)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91299158; PubMed=2069717;
RA      Redecker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;
RT      "Molecular organization of the human cathepsin D gene.";
RL      DNA Cell Biol. 10:423-431(1991).
RN      (5)
RP      SEQUENCE OF 1-22 FROM N.A.
RX      MEDLINE=94085791; PubMed=8262386;
RA      May F.E., Smith D.J., Westley B.R.;
RT      "The human cathepsin D-encoding gene is transcribed from an estrogen-
RL      regulated and a constitutive start point.";
RN      (6)
RP      Gene 134:277-282(1993).
RN      (7)
RP      SEQUENCE OF 1-22 FROM N.A.
RX      MEDLINE=95021301; PubMed=7935485;
RA      Augereau P., Mitalles F., Cavailles V., Gaudelot C., Parker M.,
RT      Rochelet H.;
RL      "Characterization of the proximal estrogen-responsive element of
RL      human cathepsin D gene.";
RN      (8)
RP      Mol. Endocrinol. 8:693-703(1994).
RN      (9)
RP      SEQUENCE OF 170-180.
RX      TISSUE=Liver;
RA      Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravlier F.,
RT      Pasquall C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
RL      Appel R.D., Hughes G.J.;
RN      Submitted (JUN-1992) to the SWISS-PROT data bank.
RN      (10)
RP      VARIANT VAL-58.
RX      MEDLINE=20179010; PubMed=10716266;
RA      Papasotiriopoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,
RT      Maier W., Pauls J., Lautenschlager N., Heun R.;
RL      "A genetic variation of cathepsin D is a major risk factor for
RN      Alzheimer's disease.";
RN      Ann. Neurol. 47:399-403(2000).
RN      (11)
RP      X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RX      TISSUE=Spleen;

```

---

```

RX      MEDLINE=93223670; PubMed=8467789;
RA      Metcalfe P., Fusek M.;
RT      "Two crystal structures for cathepsin D: the lysosomal targeting
RL      signal and active site.";
RN      EMBO J. 12:1293-1302(1993).
RN      (12)
RP      X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX      TISSUE=Liver;
RA      MEDLINE=93342076; PubMed=8393577;
RT      Balawin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,
RL      Cachau R.E., Collins J., Silva A.M., Erickson J.W.;
RN      "Crystal structures of native and inhibited forms of human cathepsin
RL      D: implications for lysosomal targeting and drug design.";
RN      Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).
CC      -1- FUNCTION: CATHEPSIN D IS AN ACID PROTEASE ACTIVE IN INTRACELLULAR
CC      PROTEIN BREAKDOWN.
CC      -1- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
CC      that of pepsin A. Does not cleave the 4-Gln-1-His-5 bond in B
CC      chain of insulin.
CC      -1- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
CC      -1- POLYMORPHISM: The Val-58 allele is significantly overrepresented
CC      in demented patients (11.8%) compared with nondemented controls
CC      (4.9%). Carriers of the Val-58 allele have a 3.1-fold increased
CC      risk for developing AD than noncarriers.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC      EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M11233; AAB59529.1; -
DR      EMBL; X05344; CAA28955.1; -
DR      EMBL; M63138; AAA51922.1; -
DR      EMBL; M63134; AAA51922.1; JOINED.
DR      EMBL; M63135; AAA51922.1; JOINED.
DR      EMBL; M63136; AAA51922.1; JOINED.
DR      EMBL; M63137; AAA51922.1; JOINED.
DR      EMBL; L12980; AAA16314.1; -
DR      EMBL; S74689; AAD14156.1; -
DR      EMBL; S52557; AAD13868.1; -
DR      PIR; A25771; KKHUDD.
DR      PDB; 1LYA; 3I-JAN-94.
DR      PDB; 1LYB; 3I-JAN-94.
DR      MEROPS; A01.009; -
DR      SWISS-2DPAGE; P07339; HUMAN.
DR      Siens-2DPAGE; P07339; -
DR      MIM; 116840; -
DR      InterPro; IPR001969; ASP_protease.
DR      InterPro; IPR001461; Pepsin.
DR      Pfam; PF00026; asp; 1.
DR      PRINTS; PR00792; PEPsin.
DR      PROSITE; PS00141; ASP_PROTEASE; 2.
KW      Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Signal; Zymogen;
KW      Polymorphism; Alzheimer's disease; 3D-structure.
FT      SIGNAL 1 18
FT      PROPEP 19 64
FT      CHAIN 65 412
FT      CHAIN 65 161
FT      CHAIN 169 412
FT      ACT_SITE 97 97
FT      ACT_SITE 295 295
FT      DISULFID 91 160
FT      DISULFID 110 117
FT      DISULFID 286 290
FT      DISULFID 329 366
FT      CARBOHYD 134 134
FT      CARBOHYD 263 263

```

N-LINKED (GLCNAC. . .)  
N-LINKED (GLCNAC. . .)

FT	VARIANT	58	58	A -> V (ASSOCIATED WITH INCREASED RISK IN AD; POSSIBLY INFLUENCES SECRETION AND INTRACELLULAR MATURATION; IN DBSNP:17571).
FT				/FTID=VAR_011621.
FT	STRAND	67	74	
FT	TURN	75	77	
FT	STRAND	78	85	
FT	TURN	86	89	
FT	STRAND	90	97	
FT	TURN	98	99	
FT	STRAND	103	107	
FT	TURN	108	109	
FT	STRAND	112	113	
FT	TURN	115	118	
FT	STRAND	119	123	
FT	TURN	123	123	
FT	STRAND	125	127	
FT	TURN	129	130	
FT	STRAND	132	141	
FT	TURN	146	158	
FT	STRAND	172	184	
FT	TURN	188	192	
FT	STRAND	197	200	
FT	TURN	204	206	
FT	STRAND	208	210	
FT	TURN	214	220	
FT	STRAND	221	222	
FT	TURN	223	233	
FT	STRAND	243	247	
FT	TURN	248	248	
FT	STRAND	252	253	
FT	TURN	255	263	
FT	STRAND	267	267	
FT	TURN	268	269	
FT	STRAND	270	279	
FT	TURN	280	281	
FT	STRAND	284	285	
FT	TURN	287	288	
FT	STRAND	290	290	
FT	TURN	296	297	
FT	STRAND	298	298	
FT	TURN	301	303	
FT	STRAND	305	315	
FT	TURN	316	316	
FT	STRAND	318	319	
FT	TURN	322	323	
FT	STRAND	325	328	
FT	TURN	329	334	
FT	STRAND	338	344	
FT	TURN	343	344	
FT	STRAND	345	349	
FT	TURN	351	354	
FT	STRAND	355	357	
FT	TURN	359	362	
FT	STRAND	365	366	
FT	TURN	370	372	
FT	STRAND	377	379	
FT	TURN	383	385	
FT	STRAND	387	390	
FT	TURN	391	392	
FT	STRAND	393	398	
FT	TURN	399	402	
FT	STRAND	403	409	

Query Match 13.7%; Score 295.5; DB 1; Length 412;  
Best Local Similarity 28.5%; Pred. No. 4.2e-17;  
Matches 105; Conservative 58; Mismatches 146; Indels 59; Gaps 15;

QY 30 YVEMTVGSPQTLNIIIVDQSSNFVGAAPHPL-----HRYYORQLSSTYRDLRKGV 83  
DB 79 YGEGIGTIPQCTTVVFDIGSSNLMVPSIHCKLIDLACWTHHKYNDKSKSTYVKNKTSF 138

QY	84	YVPYQKMEGELGTLVLSIP-----HGPNTVIRANIAITESDKFFLINGSNMEGI 134
DB	139	DIHGGSSLSGLVSDIVSPQCOSASSASALGVKXERQVFGATNQPGITFLIAAFDCI 198
QY	135	LGTAIEIARPPDLSLEFPDLSLVKQTHV-PNFSLOLCAGFPLNOSVYASVGSMTIG 193
DB	199	LGMAVPRIS--VNNVLEPVFNLMQOKLVDQNIFFSYL-----SRDDPAQGGELMIG 248
QY	194	GIDHSLYTGLMTPTIREKVEYEVIVREI-NGODLKNCKREYNKDKSIVDSGTTNLR 252
DB	249	GTDKYYKGSLSLYNTRKAWOVHLDQVEVAGSLTL---CKE--GCEALVDYGTSLMVG 303
QY	253	PKRYEAAVRSIKAASTEFKPDGFVLGEOLV-CWQAGTTPMNIIPVLSLYLGEVYTNOS 311
DB	304	PVDEVRELQAKAGVPLIQ-----GEYMIPOEKVST-----LPAITLKLGG----KG 346
QY	312	FRITILPQOYLRPEVDVATSDQCYKFAISQ-----SSTGVGAVIMEEFYVVPDRK 366
DB	347	YKLS--PEDYTLKVSQAKTL--CLSGFMGMDIPPPSGPLMLIGDVFIRGYVYVFDKNN 402
QY	367	RIGFAVSA 374
DB	403	RVGFAEAA 410

RESULT 15  
PEP4\_RABIT STANDARD; PRT; 387 AA.  
ID PEP4\_RABIT  
AC P28713;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Pepsin II-4 precursor (EC 3.4.23.1) (Pepsin A).  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91009127; PubMed=2129536;  
RA Kageyama T., Tanabe K., Koitani O.;  
RT "Structure and development of rabbit pepsinogens. Stage-specific  
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and  
RT gene expression during development."  
RL J. Biol. Chem. 265:17031-17038(1990).  
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY, ALTHOUGH BONDS  
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE  
CC ALSO CLEAVED TO SOME EXTENT.  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.  
CC -!- DEVELOPMENTAL STAGE: PEPINOGENS IN GROUP I, II, AND III WHERE  
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE  
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPINOGEN GENES IS REGULATED BY  
CC HORMONES AND RELATED SUBSTANCES.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR PIR: D38302; D38302.  
DR HSSP; P00790; IFSN.  
DR -MEROPS; A01.001; .  
DR InterPro: IPR001969; Asp-protease.  
DR InterPro: IPR001461; Pepsin.  
DR Pfam: PF000026; asp. 1.  
DR PRINTS; PR00792; Pepsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
KW Hydroxylase; Aspartyl protease; Digestion; Stomach; Gastric juice;  
KW zymogen; signal; phosphorylation; Multigene family.  
FT SIGNAL 1 59  
FT PROPEP 16 59 ACTIVATION PEPTIDE.  
FT CHAIN 60 387 PEPsin II-4.  
FT MOD\_RES 129 129 PHOSPHORYLATION (POTENTIAL).  
FT ACT\_SITE 93 93 BY SIMILARITY.  
FT ACT\_SITE 276 276 BY SIMILARITY.  
FT DISULFID 106 111 BY SIMILARITY.  
FT DISULFID 267 271 BY SIMILARITY.

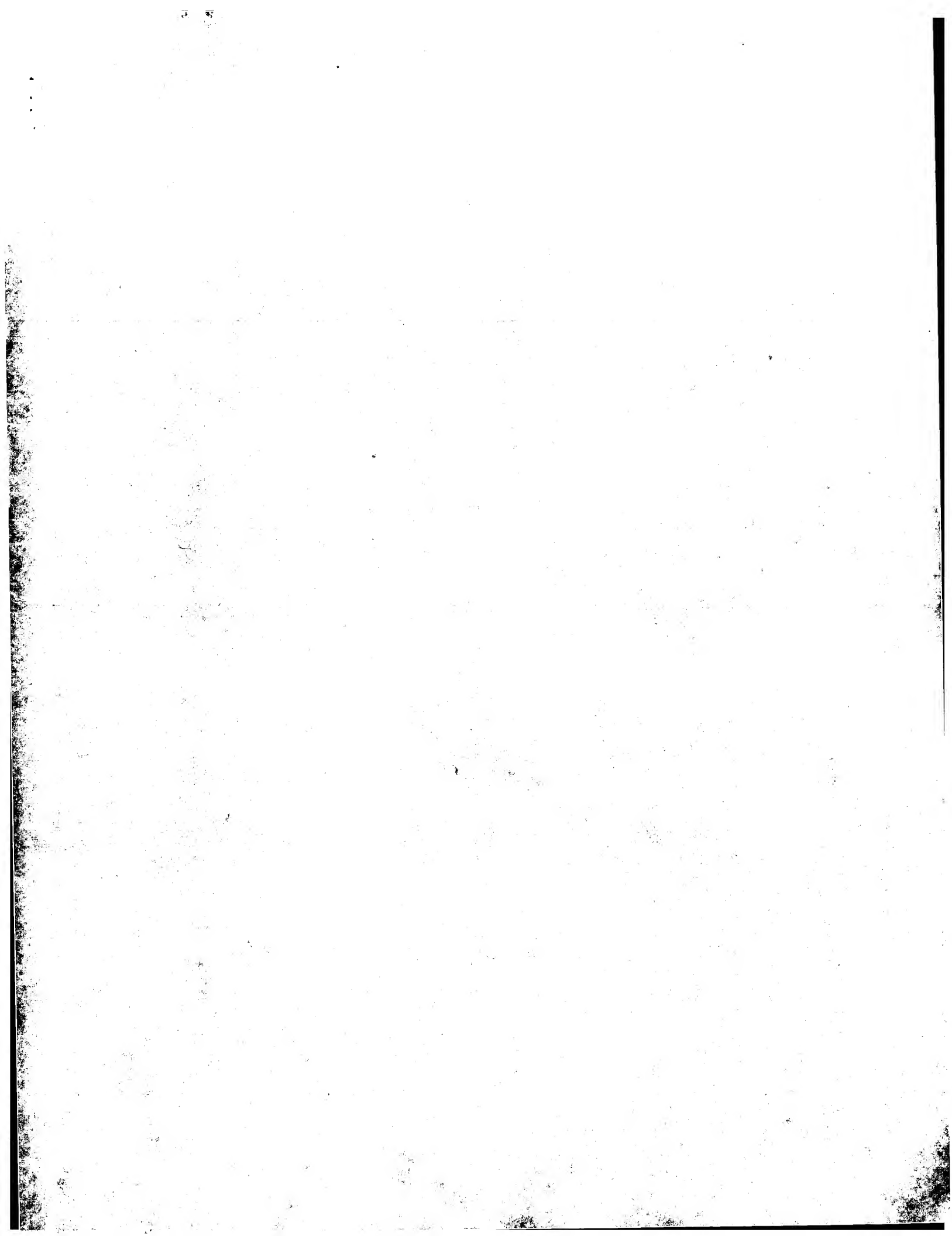
FT DISULFD 310 343 BY SIMILARITY.  
SQ SEQUENCE 387 AA; 42052 MM; 21ADD07782A89585 CRC64;

Query Match 13.7%; Score 295; DB 1; Length 387;  
Best Local Similarity 26.1%; Pred. No. 4.3e-17;

Matches 97; Conservative 65; Mismatches 123; Indels 86; Gaps 14;

```
QY 30 YVVENTVSSPPQTNILVDTGSSNFVAGAAPHP-----LHRYQROLSTYRDLKRG 82
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 75 YEGTISIGTPPDFTVLEFDIGSSNLWV---PSTYCSSLACALHRRFNPEDDSTYQGTSEF 131
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 83 YVVPYTOCKWEGELGTDLVSIPIHGPNTVRRANIAITESDKFF-----INGSNWE 132
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 132 LSTYGTGSMTGILGYDPTV-----KVGSIEDTNGIFGLSKTEPGLTFEPAPFD 179
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 133 GILGLAVAEIARPPDLSLEPPFDSLVKQTHV-PNLFSLQLCGAGFPPLNQSEVLASVGGSMT 191
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 180 GILGLAVPSISSDPAF--PVFDNMNMNEGILVSQDLFSVYLSSDD-----EKGSLVM 227
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 192 IGGIDHSLYTGSLWYFPFIREWYEVYIIVREINGODLKM--DCKEYNDKSTIVDSGTN 249
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 228 FGGIDSSYTTGSLWVWVSYEGYWOITMDSVINGETIACADSC-----QATVDGTSL 281
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 250 LRLPKKYFEAAVKSIRKASSTEFPPDGFVLGEQLV-CWQAGTTPMNIIPVYISLYLMEGYV 308
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 282 LTGP---TSAISNIOISYIGASK---NLGENVVISCSAIDSLPDIYF----- 321
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 309 NOSFRITILPQOYLRPVEDVATSQDDCYKFAISQSTGT-----VMGAVIMEGYVVD 362
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 322 -----TINCIOYPLPASAIILKEDDCTSGLEGMNVDTYTGELWILGDVFIROYPTVFD 375
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 363 RARRIGFAVS 373
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 376 RANNOGLGLAA 386
```

Search completed: August 7, 2002, 09:15:12  
Job time: 96 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2002, 09:17:12 ; Search time 50.29 Seconds

(without alignments)  
1400.058 Million cell updates/sec

Title: US-09-724-571-58

Perfect score: 2156

Sequence: 1 ETDEPEEPGRGSGFEVEMVD.....GPFVTLDMEDCGYNIPQIDE 407

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2156	100.0	532	4	Q9ULS1
2	2001.5	92.8	476	4	Q9BYC1
3	1898	88.0	457	4	Q9BYC0
4	1753.5	81.3	432	4	Q9BYB9
5	1149	53.3	266	11	Q9CUU5
6	1121	52.0	439	4	Q9H2V8
7	1116.5	51.8	514	11	Q9JL18
8	967.5	44.9	396	4	Q9NZL1
9	933	43.3	468	4	Q9NZL2
10	712.5	33.0	213	4	Q9P0D2
11	566.5	26.3	253	11	Q9R1P7
12	332.5	15.4	391	5	Q9VKP6
13	332	15.4	354	5	Q9GYX7
14	312.5	14.5	386	6	Q9BGU5
15	308	14.3	384	13	Q9DEC2
16	308	14.3	385	13	Q9DEC4

17	305	14.1	386	6	Q9GMV7	Q9gmv7 rhinolophus
18	305	14.1	387	6	Q9GMV8	Q9gmv8 soxex ungu
19	304.5	14.1	372	5	Q9YIK3	Q9yik3 drosophila
20	304.5	14.1	387	13	Q9DDV5	Q9ddv5 salvelinus
21	304	14.1	387	6	Q9GMV9	Q9gmv9 suncus murt
22	303.5	14.1	383	13	Q9DE45	Q9de45 salvelinus
23	303	14.1	383	13	Q9DEC3	Q9dec3 xenopus lae
24	302.5	14.0	376	13	Q9PUR8	Q9pur8 pseudopleur
25	301	14.0	382	13	Q9PRG9	Q9prg9 gallus gall
26	298.5	13.8	423	5	Q9YKP7	Q9ykp7 drosophila
27	298	13.8	384	13	Q9I3J2	Q9i3j2 rana catesb
28	295.5	13.7	386	6	Q9GMV6	Q9gmv6 canis fami
29	294	13.6	396	13	Q934J8	Q934j8 chionodraco
30	290.5	13.5	381	6	Q9GK11	Q9gk11 camelus dro
31	288	13.4	399	13	Q934J8	Q934j8 podarcis si
32	287.5	13.3	444	5	Q21966	Q21966 caenorhabdi
33	284	13.2	398	13	P87370	P87370 oncorhynch
34	284	13.2	427	5	P91802	P91802 schistosoma
35	281	13.0	378	13	Q9PUR9	Q9pur9 pseudopleur
36	279.5	13.0	390	6	Q9GK10	Q9gk10 camelus dro
37	278	12.9	370	6	Q9TWT1	Q9twt1 bos taurus
38	278	12.9	399	13	Q9D89	Q9d89 brachydanio
39	278	12.9	422	5	Q96906	Q96906 onchocerca
40	277.5	12.9	380	6	Q28950	Q28950 sus scrofa
41	277	12.8	446	5	Q9N9H3	Q9n9h3 neocator ame
42	276	12.8	389	13	Q9PWK1	Q9pwk1 gallus gall
43	276	12.8	396	13	Q9W643	Q9w643 gallus gall
44	276	12.8	396	13	Q9DEX3	Q9dex3 clupea hare
45	275	12.8	385	6	Q29080	Q29080 sus scrofa

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	532 AA.
1	Q9ULS1			
AC	Q9ULS1			
DT	01-MAY-2000 (TRENDArel. 13, Created)			
DT	01-OCT-2001 (TRENDArel. 18, Last sequence update)			
DT	01-DEC-2001 (TRENDArel. 19, Last annotation update)			
DE	KIAA1149 PROTEIN (FRAGMENT).			
GN	KIAA1149.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RX	MEDLINE=20039618; PubMed=10574461;			
RA	Hiroseawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.,			
RT	*Characterization of cDNA clones selected by the Genemark analysis			
RT	from size-fractionated cDNA libraries from human brain."			
RL	DNA Res. 6:329-336(1999).			
DR	EMBL; AB032975; BAA8463.2; -			
DR	HSSP; P56272; IAMS.			
DR	MEROPS; A01.004; -			
DR	InterPro; IPR001960; ASP-protease.			
DR	Pfam; PF00026; asp. 1.			
DR	PRINTS; PR00792; PEPsin.			
DR	PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.			
FT	NON_TER			
SO	SEQUENCE 532 AA; 58720 MW; 98B135D05FBD2E8 CRC64;			

Query Match 100.0%; Score 2156; DB 4; Length 532;  
Best Local Similarity 100.0%; Pred. No. 3.3e-175;  
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ETDEPEEPGRGSGFEVEMVDNLKSGGCIYVEMVGSPPQTLNLVDTGSSNFAVGAAP 60  
|||||

```

Db 77 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAP 136
QY 61 HPFLHRYRQRLSTYRDLRKGVVPTTQCKNKEGLGTDLVSIIPHGNVTVRANIAATE 120
Db 137 HPFLHRYRQRLSTYRDLRKGVVPTTQCKNKEGLGTDLVSIIPHGNVTVRANIAATE 196
QY 121 SDFPFINSNMWEGILGLAVAEIARPDSDLPEFPDLSVKQTHVPMNLFSLQCGAGFPLNOS 180
Db 197 SDFPFINSNMWEGILGLAVAEIARPDSDLPEFPDLSVKQTHVPMNLFSLQCGAGFPLNOS 256
QY 181 EYLVASVGSMTIGGIDHSLYTGSLMTYPIRREMYEVIIVVEINQDLKMDCKEYNDK 240
Db 257 EYLVASVGSMTIGGIDHSLYTGSLMTYPIRREMYEVIIVVEINQDLKMDCKEYNDK 316
QY 241 SYVDSGTTNLRPKKVEFAVAKSIRKAASTKEKPDGFMGLGDLVCMQAGTTPMNIFFPIS 300
Db 317 SYVDSGTTNLRPKKVEFAVAKSIRKAASTKEKPDGFMGLGDLVCMQAGTTPMNIFFPIS 376
QY 301 LYLMGEVTVNOSFRITILLPOQYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV 360
Db 377 LYLMGEVTVNOSFRITILLPOQYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV 436
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMDCGYNIPQIDE 407
Db 437 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMDCGYNIPQIDE 483

RESULT 2
Q9BYC1 PRELIMINARY: PRT: 476 AA.
ID 09BYC1
AC 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-476.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RA MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production."
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050436; BAB40931.1; -.
DR HSSP; P32329; IYPS.
DR InterPro; IPR001969; ASP_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 476 AA; 52907 MW; 6C8C87F8A953AF66 CRC64;

```

```

Query Match 92.8%; Score 2001.5; DB 4; Length 476;
Best Local Similarity 93.9%; Pred. No. 4; Le-162; Indels 25; Gaps 1;
Matches 382; Conservative 0; Mismatches 0;

```

```

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAP 60
Db 46 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAP 105
QY 61 HPFLHRYRQRLSTYRDLRKGVVPTTQCKNKEGLGTDLVSIIPHGNVTVRANIAATE 120
Db 106 HPFLHRYRQRLSTYRDLRKGVVPTTQCKNKEGLGTDLVSIIPHGNVTVRANIAATE 165
QY 121 SDFPFINSNMWEGILGLAVAEIARPDSDLPEFPDLSVKQTHVPMNLFSLQCGAGFPLNOS 180

```

```

Db 166 SDFPFINSNMWEGILGLAVAEIARPDSDLPEFPDLSVKQTHVPMNLFSLQCGAGFPLNOS 200
QY 181 EYLVASVGSMTIGGIDHSLYTGSLMTYPIRREMYEVIIVVEINQDLKMDCKEYNDK 240
Db 201 EYLVASVGSMTIGGIDHSLYTGSLMTYPIRREMYEVIIVVEINQDLKMDCKEYNDK 260
QY 241 SYVDSGTTNLRPKKVEFAVAKSIRKAASTKEKPDGFMGLGDLVCMQAGTTPMNIFFPIS 300
Db 261 SYVDSGTTNLRPKKVEFAVAKSIRKAASTKEKPDGFMGLGDLVCMQAGTTPMNIFFPIS 320
QY 301 LYLMGEVTVNOSFRITILLPOQYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV 360
Db 321 LYLMGEVTVNOSFRITILLPOQYLRPVEDVATSDDCYKFAISOSTGTVMGAVIMEGFYV 380
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMDCGYNIPQIDE 407
Db 381 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMDCGYNIPQIDE 427

```

```

RESULT 3
Q9BYC0 PRELIMINARY: PRT: 457 AA.
ID 09BYC0
AC 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-457 (BETA-SITE APP CLEAVING ENZYME
DE TYPE C).
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RA MEDLINE=21408467; PubMed=11516562;
RA Tanahashi H., Tabira T.;
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production."
RL Neurosci. Lett. 307:9-12(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EXOCRINE PANCREAS;
RC Zaccarelli D., De Pietri Tonelli D., Schnurbus R.;
RA "New beta-site APP cleaving enzyme isoform (BACE-1C) obtained from
RA human pancreas."
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050437; BAB40932.1; -.
DR EMBL; AF338817; AAK38375.1; -.
DR HSSP; P32329; IYPS.
DR InterPro; IPR001969; ASP_protease.
DR InterPro; IPR001461; Pepsin.
DR Pfam; PF00026; asp. 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 457 AA; 51068 MW; C794C9A985FE7A2 CRC64;

```

```

Query Match 88.0%; Score 1898; DB 4; Length 457;
Best Local Similarity 89.2%; Pred. No. 2; Se-153; Indels 44; Gaps 1;
Matches 363; Conservative 0; Mismatches 0;

```

```

QY 1 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAP 60
Db 46 ETDEPEEPGRGSGFVEMVDNLKSGGGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAP 105
QY 61 HPFLHRYRQRLSTYRDLRKGVVPTTQCKNKEGLGTDLVSIIPHGNVTVRANIAATE 120
Db 106 HPFLHRYRQRLSTYRDLRKGVVPTTQCKNKEGLGTDLVSIIPHGNVTVRANIAATE 145
QY 121 SDFPFINSNMWEGILGLAVAEIARPDSDLPEFPDLSVKQTHVPMNLFSLQCGAGFPLNOS 180

```

```
Db 146 -----PDDSLPEFDSLVKQTHVNPFLSLQCGAGFPLNDS 181
QY 181 EVLASVGSMTIGIDHSILYTGSLMTPIRREMYEVIIVRAEINODLKMCKEYNDK 240
Db 182 EVLASVGSMTIGIDHSILYTGSLMTPIRREMYEVIIVRAEINODLKMCKEYNDK 241
QY 241 SYDSGTTNLRPKKVEAAVKSIIKAASTKPPDGFMLGEOLVCMQAGTTPMNIIPVIS 300
Db 242 SYDSGTTNLRPKKVEAAVKSIIKAASTKPPDGFMLGEOLVCMQAGTTPMNIIPVIS 301
QY 301 LYLMGEVTVNSFRITILLPOOYLRPVEDVATSODDCYKFAISOSSTGTGAVIMGEFYV 360
Db 302 LYLMGEVTVNSFRITILLPOOYLRPVEDVATSODDCYKFAISOSSTGTGAVIMGEFYV 361
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTE 407
Db 362 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTE 408

RESULT 4
Q9BYB9 PRELIMINARY; PRT; 432 AA.
AC Q9BYB9:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME I-432.
GN BACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID:9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-21408467; PubMed-11516562;
RA Tanahashi H., Tabira T.,
RT "Three novel alternatively spliced isoforms of the human beta-site APP
RT cleaving enzyme (BACE) and their effect on amyloid beta-peptide
RT production."
RL Neurosci. Lett. 307:9-12(2001).
DR EMBL; AB050438; BAB40933.1; -.
DR HSSP; P32329; IYPS.
DR InterPro: IPR001969; Asp_protease.
DR Pfam: PF00026; asp. 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 432 AA; 48212 MW; 96FC01E6F0ED01B CRC64;
```

```
Query Match 81.3%; Score 1753.5; DB 4; Length 432;
Best Local Similarity 83.0%; Pred. No. 4.8e-141;
Matches 338; Conservative 0; Mismatches 0; Indels 69; Gaps 1;
```

```
QY 1 ETDEPEEPGRGSRGVENVNDLRKSGGGYVEMTVGSPPTLILIVDTGSSNFAVGAAP 60
Db 46 ETDEPEEPGRGSRGVENVNDLRKSGGGYVEMTVGSPPTLILIVDTGSSNFAVGAAP 105
QY 61 HPFLHRYQROLSTYRDLRKGVVVPYTGKMEGELGTDLSIPHGNVYRANIAITE 120
Db 106 HPFLHRYQROLSTYRDLRKGVVVPYTGKMEGELGTDLSIPHGNVYRANIAITE 145
QY 121 SDRFFINSMNEGILGLAYEAIARPDDSLVFKQTHVNPFLSLQCGAGFPLNDS 180
Db 146 -----LCGAGFPLNDS 156
QY 181 EVLASVGSMTIGIDHSILYTGSLMTPIRREMYEVIIVRAEINODLKMCKEYNDK 240
Db 157 EVLASVGSMTIGIDHSILYTGSLMTPIRREMYEVIIVRAEINODLKMCKEYNDK 246
QY 241 SYDSGTTNLRPKKVEAAVKSIIKAASTKPPDGFMLGEOLVCMQAGTTPMNIIPVIS 300
```

```
Db 217 SYDSGTTNLRPKKVEAAVKSIIKAASTKPPDGFMLGEOLVCMQAGTTPMNIIPVIS 276
QY 301 LYLMGEVTVNSFRITILLPOOYLRPVEDVATSODDCYKFAISOSSTGTGAVIMGEFYV 360
Db 277 LYLMGEVTVNSFRITILLPOOYLRPVEDVATSODDCYKFAISOSSTGTGAVIMGEFYV 366
QY 361 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTE 407
Db 337 FDRARRRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTE 383
```

```
RESULT 5
Q9CUC5 PRELIMINARY; PRT; 266 AA.
AC Q9CUC5:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-SITE APP CLEAVING ENZYME (FRAGMENT).
GN BACE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-BRAIN;
RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Cassavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochila H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK014390; BAB29317.1; -.
DR MEROPS; A01.004; -.
DR MGD; MGI:1346542; Bace.
DR InterPro: IPR001461; Pepsin.
DR Pfam; PF00026; asp. 2.
FT NON_TER 1
SQ SEQUENCE 266 AA; 30201 MW; B913FDA8ADAB4238 CRC64;
```

```
Query Match 53.3%; Score 1149; DB 11; Length 266;
Best Local Similarity 99.1%; Pred. No. 7.7e-90;
Matches 215; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 191 IIGIDHSILYTGSLMTPIRREMYEVIIVRAEINODLKMCKEYNDKSYDSGTTNLR 250
Db 1 IIGIDHSILYTGSLMTPIRREMYEVIIVRAEINODLKMCKEYNDKSYDSGTTNLR 60
QY 251 RLKPKVFEAAVKSIIKAASTKPPDGFMLGEOLVCMQAGTTPMNIIPVISLYLMGEVTV 310
Db 61 RLKPKVFEAAVKSIIKAASTKPPDGFMLGEOLVCMQAGTTPMNIIPVISLYLMGEVTV 120
QY 311 SPRTITLPOOYLRPVEDVATSODDCYKFAISOSSTGTGAVIMGEFYVDFDRARRRIGF 370
Db 121 SPRTITLPOOYLRPVEDVATSODDCYKFAISOSSTGTGAVIMGEFYVDFDRARRRIGF 180
```

RESULT	7	
Q9JL18		
ID	Q9JL18	PRELIMINARY;
AC	Q9JL18;	PRT; 514 AA.
DT	01-OCT-2000 (TrEMBLrel, 15, Created)	
DT	01-OCT-2000 (TrEMBLrel, 15, Last sequence update)	

05 Homo sapiens (Human).  
06 .....  
0C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
0C Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
0C .....  
0X NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE-20422477; PubMed-10965118;  
RA Solans A., Estivill X., de La Luna S.;  
RT "A new aspartyl protease on 21q22.3, BACE2, is highly similar to  
RL Alzheimer's amyloid precursor protein beta-secretase.";  
RL Cyogenet. Cell Genet. 89:177-184(2000).  
DR EMBL; AF188277; AAF35836.1; -  
DR HSSP; P00797; ZREN.  
DR MEROPS; A01.041; -  
DR InterPro; IPR001969; Asp\_protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; Asp\_2.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_2.  
KW Protease.  
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

Query Match	44.98;	Score 967.5;	DB 4;	Length 396;
Best Local Similarity	54.68;	Pred. No. 4e-74;		
Matches 183;	Conservative	54;	Mismatches	85;
			Indels	13;
			Gaps	4;

[illegible]

RESULT	9	Q9NZL2	Q9NZL2
ID	Q9NZL2	PRELIMINARY;	PRT; 468 AA.
AC	Q9NZL2;		
DT	01-OCT-2000 (TREMblrel, 15, Created)		
DT	01-OCT-2000 (TREMblrel, 15, Last sequence update)		
DT	01-DEC-2001 (TREMblrel, 19, Last annotation update)		
DE	ASPARTYL PROTEASE.		
GN	BACE2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
NC	NCBI_TaxID=9606;		
NR	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20422477; PubMed-10965118;		
RA	Soizans A., Estivill X., de La Luna S.;		
RT	"A new aspartyl protease on 21q22.3, BACE2, is highly similar to		
RL	Alzheimer's amyloid precursor protein beta-secretase.";		
RL	Cytogenet. Cell Genet. 89:177-184(2000).		
DR	EMBL: AF188276; AAF35635.1; .		
DR	HSSP: P00797; 2REN.		
DR	MEROPS; A01.041; .		
DR	InterPro: IPR001969; Asp--protease.		
DR	InterPro: IPR001461; Pepsin.		
DR	Pfam: PF000026; asp; 3.		

DR PRINTS: PRO0792; PEPSIN.  
DR PROSITE; PS00141; ASP\_PROTEASE; UNKNOWN\_2.  
KW Protease.  
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64

Query Match	43.3%	Score 933;	DB 4;	Length 468;
Best Local Similarity	46.5%;	Pred. NO. 4.5e-71;		
Matches 180;	Conservative 59;	Mismatches 94;	Indels 54;	Gaps 3

```

QY          9 PGAGSGVEWVDLNRKSGOGYVEVTVGSPPQTLNLLVDTGSSNFAVGAAPHPFLHRY 68
Db          71 PAGANLANVAWLDGDSRGYLLLEMLIGTPQKQIILVDTGSSNFAVAGPHSIDTYF 130
QY          69 QRLSSTRYRDLRFKGVVVPYTOGKMEGELGTDVSLPHGPNTVYANALATSESKPPING 128
Db          131 DTERSSYTRKSGDYVKTQGSWTFGEVGEDLVITIPKGFNTSPLVNATTFESBNFLFG 190
QY          129 SNNEGILGAYAEIARPDLSLEPFDSIVKQTHVNPILFSLDLCAGPFLNOSWEVLASVVG 188
Db          191 IKWNGILGAYATLAPKSSLETFPDSLVTQANIPNPNVSMCMCAGLPVAGS---GTNGG 247
QY          189 SMIIIGIDISLVTSGLMTYPIPIREKRYEYIVTVRVLNGIDKIMDKCKEYNDKSIDVSGTT 248
Db          248 SLVPGIGIESLTKGDIVWYPIKMEWYIOETILKEIGSGSLNDPCREYNADKAIYDSGTT 307
QY          249 NLRLPKVEEAVKSTIKASSTREKPPDOEFWLGEBDLYCMQAGTTPMNLFPVLSLYLMEGYT 308
Db          308 LRLPKVEEAVENAVARSL----- 329
QY          309 NOSFRITLLPOOYLRPVEDVATISODDCYKFAISOSTGTVMGAVIMEGFVYVDRARKRI 368
Db          330 -----YIOPMKGAGLNT-ECYRFGESISTNALVIGATVMEGFYVIFDRQKRY 376
QY          369 GFAVSACHHDEFRTAAVEGPFVLDM 395
Db          377 GRASPCEALIGAAVSEISGPFSTEDV 403

```

RESULT		10	
ID	Q9P0D2	PRELIMINARY;	PRT; 213 AA.
AC	Q9P0D2;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	HSPc104 (FRAGMENT).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
NCBI_OX	NCBI_TaxId:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	TISSUE=CORD BLOOD;		
RR	Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.O., Wang L., Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.,		
RR	Human partial CDS cloned from cd34+ stem cells.;		
RR	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.		
RR	EMBL; AF161367; AAF28927.1; -		
RR	Interpro: IPR001461; Pepsin.		
RR	Pfam: PF00026; asp; 1.		
FT	NON TER		
FT	1		
SSO	SEQUENCE	213 AA;	24338 MW; EC9D3FPA31CEFA835C CRC64;

Query Match	33.0%	Score 712.5;	DB 4;	Length 213;
Best Local Similarity	83.58;	Prid. No. 8.7e-53;		
Matches 137;	Conservative 4;	Mismatches 12;	Indels 11;	Gaps 1;

193	GGIDHSLTGSLMTPTIRREYEVIVILVREVLINGODIKMKCKEYNDKSDIVSGTNNRL	252
1	GGIDHSLTGSLMTPTIRREYEVIVILVREVLINGODIKMKCKEYNDKSDIVSGTNNRL	60

[illegible]

NCBI\_TaxID=1227;  
 11] ]]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-BERKELEY;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer V.G., Chapple M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abvill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu P.V., Bernard B.P., Bhandari D., Bolshakov S.,  
 RA Beeson K.Y., Bemis P.V., Bertone P., Brokstein P., Brotlier P.,  
 RA Borokova D., Botchan M.R., Bouck H., Cadieu E., Center A., Chandra I.,  
 RA Butts K.C., Busam D.A., Butler H., Cayleport L.B., Davies P.,  
 RA Cherry J.M., Cavley S., Dahlke K., Davenport L.B., Davies S.M.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinelte K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spirdlian A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weissensbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 CC "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1, ALSO KNOWN AS THE  
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
 DR EMBL: AE003630; AAF53016.1; -  
 DR HSSP: P00794; 4CMS.  
 DR MEROPS: A01.0PM; -  
 DR Flybase: FBgn0032304; CG17134.  
 DR InterPro: IPR001969; Asp-protease.  
 DR InterPro: IPR001461; Pepsin.  
 DR Pfam: PF00026; asp; 1.  
 DR PRINTS: PRO0792; PEPSIN.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 2.  
 KW Aspartyl protease: Hydrolase.  
 SEQUENCE 391 AA; 42016 MW; 99A18E7131025E1B CRC64;

QY 193 GGDHSLTGTGSLMTYPIREWEYEVIVVEINGDLMKDEKEYNYSIYSGTINRL 252  
DB 234 GGDHSLTGTGSLMTYPIREWEYEVIVVEINGDLMKDEKEYNYSIYSGTINRL 287  
QY 253 PKKVEAAVKSIIKASSTEEKFPDGLWGLGOLY-CWQAGTTTWNIPVYSILYLMGEVYNOS 311  
DB 268 PLAVARKINROIGATDND-----GEAFVRCGRVSS-----LPKVNINIGSTV----- 329  
QY 312 FRITLPOOYLPRVEDVATSDDDCYK-FAISOSSTGYMGAVIMEGFYVYVDFRARRKIGF 370  
DB 330 --FTLAPRDYI--VKVQNGOTYCMASFTYMEGLSFWLGDVFTGKFTYVDFKNGMERIGF 385  
QY 371 A 371  
DB 386 A 386

RESULT 13  
09GXX7 PRELIMINARY; PRT; 354 AA.  
AC 09GXX7; 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE HEME-BINDING ASPARTIC PROTEINASE (FRAGMENT).  
OS Boophilus microplus (cattle tick).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Boophilus.  
NCBI\_TaxID=6941;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PORTO ALEGRE; TISSUE=OVARY;  
RA Sorghum M.H.F., Logullo C., Zingali R.B., Paiva-Silva G.O.,  
RA Juliano L., Oliveira P.L.;  
RT "A heme-binding aspartic proteinase from the eggs of the hard tick  
RT Boophilus microplus";  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL; AF286865; AAG00993.1; -  
DR HSSP; P00797; 2REN.  
DR MEROPS; A01.054; -  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPsin.  
FT NON\_TER 1  
SQ SEQUENCE 354 AA; 38512 MW; B50D7C638CF27091 CRC64;

Query Match 15.4%; Score 332; DB 5; Length 354;  
Best Local Similarity 26.2%; Pred. No. 4.9e-20;  
Matches 101; Conservative 74; Mismatches 135; Indels 76; Gaps 15;  
QY 19 VNLNGK-----SCGGYVEMTVGSPPTQNLIVDTGSSNFVAGAPHPFL- 64  
DB 13 VVEINGRALGDPRIITLNNMPOFYGIITIGTPQSEFKLMDTGSSNFWV-----PSIN 67  
QY 65 -----HRYRQRLSTYRDLRKGVYVPTQGWEGELGTDLVSIPIHPNVTYRANI 115  
DB 68 CDOSMACRDHAUYDSSSTFKSGRYIRIKRSGVNVAGTISIDNVG--GPAVVTQKFE 125  
QY 116 AAITESDKFFINGSMWEGILGLAYAEIARPDSELEPFDSLAKOTHPN-LESLQLOGAG 174  
DB 126 AEMHSDGKLRNMAKYDIFGLAPPSISQ--NNQLPLEDAVAKGVRAQAVFSLYL--SK 181  
QY 175 PPLNOSSEVLASVGSMTIGIDHSLYTGSLMTYPIREWEYEVIVVEINGDLMK-DC 233  
DB 182 QPSEQN-----GGEIYGGINAGRTGAHIVYVSOAHMÖVMDINVOGTTLCVGGC 235  
QY 234 KEYNDSKIVSGTINRLPKKFEAAVKSIIKASSTEEKFPDGLWGLGOLYCWQAGTTTPW 293  
DB 236 -----PTVVDGSGNLSGP-----SADVETLNRVIGATKTAAGY-----FEVNCATI 277  
QY 294 NIPVYSILYLMGE--VTNQSFRITILPQOYLPRVEDVATSDDDCYKFAISOSSTGT-- 347

DB 278 SSLEPTIFENLNGKSFPLQGEAYTIRI-----PLTGGEGQCTRISESADASTNLM 327  
QY 348 VMGAVIMEGFYVDFRARRKIGFAVS 373  
DB 328 ILGAVFTQTYTYVDFRANKNVGFATA 353

RESULT 14  
09BGU5 PRELIMINARY; PRT; 386 AA.  
AC 09BGU5; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE CATHEPSIN D (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Higuchi M., Miyashita N., Nagamine Y., Awata T.;  
RT "Complementary DNA sequence of bovine cathepsin D."  
RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE  
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.  
DR EMBL; AB055312; BAB21620.1; -  
DR HSSP; P07339; 1LYB.  
DR MEROPS; A01.009; -  
DR InterPro; IPR001969; Asp-Protease.  
DR InterPro; IPR001461; Pepsin.  
DR Pfam; PF00026; asp; 1.  
DR PRINTS; PR00792; PEPsin.  
DR PROSITE; PS00141; ASP\_PROTEASE; 2.  
KW Aspartyl protease; Hydrolase.  
FT NON\_TER 1  
SQ SEQUENCE 386 AA; 41965 MW; 33BF0DB80585490D CRC64;

Query Match 14.5%; Score 312.5; DB 6; Length 386;  
Best Local Similarity 28.0%; Pred. No. 2.5e-18;  
Matches 113; Conservative 69; Mismatches 131; Indels 91; Gaps 20;  
QY 8 EPG-RGSEVEMDNLKRGSGGGYVEMTVGSPPTQNLIVDTGSSNFVAGAPHPFL-- 64  
DB 35 EPPAVRQGPRLPELKNYMDAQ---YGBIGIGTPPQCFVTFDGSANLWVPSIHKCLDI 91  
QY 65 -----HRYRQRLSTYRDLRKGVY--VPTQGWEGELGTDLVSIPIHPN-----VTVR 112  
DB 92 ACWTHRKRYNDSKSTY--VANGTFEDIHGSGSLSYGLQDDVYVPCNSSSSPGGVTVQ 149  
QY 113 ANI--AAITESDKFFINGSMWEGILGLAYAEIARPDSELEPFDSLAKOTHPN-PLNESIQ 169  
DB 150 RQTFGEAIKQPGVYF-AAKFGILGMAYPRIS--VNNVLPEFDNLMOQKLYDKWVFS-- 204  
QY 170 LCGAGPPLNOSSEVLASVGSMTIGIDHSLYTGSLMTYPIREWEYEVIVVEINGDLM 229  
DB 205 -----FTLNR-DEKAPQGGELMGTDGSKYRGSIMFHNVTROAWOIHMDLDV-GSSL 257  
QY 230 KMDCKEYNDKSIYDVGSTINRLPKKFEAAVKSIIKASSTEEKFPDGLWGLGOLYCWQAG 289  
DB 258 TV-CK--GGEALIVDTGSLIYGPVEVRELKAL-----G 290  
QY 290 TTPWNIFFVYSILYLMGEVYNOSFRITILPQOYLPRVEDVATSDDDCYKFAISOSSTGT- 347  
DB 291 AVP-----LIGEVNIPCEKYSLSLPÖVYVKLGKDYAXSPED-YALVYSQAGTTVC 340  
QY 348 -----VMGAVIMEGFYVDFRARRKIGFAVS 374  
DB 341 LSGPMGMDIPPGGPLMLIGDVFGIRYTYVDFRDNRVGLAEAA 384

```
RESULT 15
09DEC2 PRELIMINARY; PRT; 384 AA.
AC 09DEC2: 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PEPSINOGEN A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikuzawa M., Inokuchi T., Kobayashi K., Yasumasa S.;
RT "Molecular Cloning of Pepsinogens in Adult Xenopus laevis and Bullfrog
RT Rana catesbeiana."
RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC EMBL: AB045380; BAB20798.1; -.
CC DR HSSP: P00790; 1PSN.
CC DR MEROPS: A01.001; -.
CC DR InterPro: IPR001969; Asp.protease.
CC DR InterPro: IPR001461; Pepsin.
CC DR Pfam: PF00026; asp_1.
CC DR PRINTS: PR00792; PEPsin.
CC DR PROSITE: PS00141; ASP_PROTEASE; 1.
CC KM Aspartyl protease; Hydrolase.
CC SQ SEQUENCE 384 AA; 41888 MW; 729DD2E7953D9072 CRC64;

Query Match 14.3%; Score 308; DB 13; Length 384;
Best local Similarity 27.5%; Pred. No. 5,1e-18;
Matches 99; Conservative 60; Mismatches 133; Indels 68; Gaps 16;

OY 30 YVEMTVGSPPOQLINILVDTGSSNFVGAAPHPL-----HRYQRLSTYRDLRG 82
DB 72 YVGTISIGTPQEFYTFDGSANLWY---PSYCSQACSNHRENPQOSTFOATNTP 128
OY 83 VYVPYTGKMGELGTLVSIHPGPNYTVRANIAITESDK-FEINGSNNEGILGLAYAE 141
DB 129 VSIQYGTGSMGFLGYDPLQY--GNIQISNQMGISESEPGSLYXSPDGLGLAFPS 185
OY 142 IARPDDSLPEPFDLSIVKOTHP-NLFSIOLGAGFPLNOSVILASVGSMTIGIDHSLY 200
DB 186 IA--SSQATPFVDNMWGGGLPQNLFSYLSDDG-----QTGSYVLFGGVDNSY 233
OY 201 TGSIMVTPIRREWYEVILVEINGDL--KMDCKEYNKSIYDSGTNLRPKKVE 258
DB 234 SGSLMWVPLFAETWQITLDSVSIQVYACSQSC-----QALVDIGTSLMTGPSTPI- 286
OY 259 AAVKSIKAASTERKPPDGFLEOLV-CWQAGTTPWNIPVYISLYLMGEVTNOSFRITL 317
DB 287 ANIQNTIGASQDSN-----GQYVINCNNISNMPTIVE-----TIN 321
OY 318 POQY-LRPVEDVATSDDDCYK-FAISQSSTGT---VMGAVIMEGFYVFPDRARKRIGFA 371
DB 322 GVOYPLSPSAVYRONOGCGSGFOAMNLPNTNSGDLMLIGDVPFROYFTVFDKANNYVAIA 381
```

Search completed: August 7, 2002, 09:17:13  
Job time: 217 sec